UNITED STATES AIR FORCE

SUMMER RESEARCH PROGRAM -- 1998

SUMMER RESEARCH EXTENSION PROGRAM FINAL REPORTS

VOLUME 4
WRIGHT LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way

Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Colonel Jan Cerveny

Program Manager, RDL
Scott Licoscos

Program Administrator, RDL
Johnetta Thompson

Program Administrator, RDL
Rebecca Kelly-Clemmons

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Bolling Air Force Base

Washington, D.C.

December 1998

20010319 030

AQM01- 06- 1181

PREFACE

This volume is part of a four-volume set that summarizes the research of participants in the 1998 AFOSR Summer Research Extension Program (SREP). The current volume, Volume 1 of 5, presents the final reports of SREP participants at Armstrong Laboratory.

Reports presented in this volume are arranged alphabetically by author and are numbered consecutively -- e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3, with each series of reports preceded by a 35 page management summary. Reports in the five-volume set are organized as follows:

VOLUME	TITLE
1	Armstrong Research Laboratory
2	Phillips Research Laboratory
3	Rome Research Laboratory
4	Wright Research Laboratory
5	Air Logistics Center Arnold Engineering Development Center

REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-00-

0717

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, s the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including su Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22352-4302, and to the Office of Management and Budget, Pa

d reviewing Information

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE December, 1998	3. REP

4. TITLE AND SUBTITLE 1998 Summer Research Program (SRP), Summer Research Extension Program (SREP), Final Report, Volume 4, Wright Laboratory	5. FUNDING NUMBERS F49620-93-C-0063

6. AUTHOR(S)
Gary Moore

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Research & Development Laboratories (RDL) 5800 Uplander Way Culver City, CA 90230-6608	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research (AFOSR) 801 N. Randolph St. Arlington, VA 22203-1977	10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release	12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The United States Air Force Summer Research Program (SRP) is designed to introduce university, college, and technical institute faculty members to Air Force research. This is accomplished by the faculty members, graduate students, and high school students being selected on a nationally advertised competitive basis during the summer intersession period to perform research at Air Force Research Laboratory (AFRL) Technical Directorates and Air Force Air Logistics Centers (ALC). AFOSR also offers its research associates (faculty only) an opportunity, under the Summer Research Extension Program (SREP), to continue their AFOSR-sponsored research at their home institutions through the award of research grants. This volume consists of a listing of the participants for the SREP and the technical report from each participant working at the AF Wright Laboratory.

14. SUBJECT TERMS Air Force Research, Air Force, Engineering, Laboratories, Reports, Summer, Universities, Faculty, Graduate Student, High School Student	15. NUMBER OF PAGES
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

Standard Form 298 (Rev. 2-89) (EG)
Prescribed by ANSI Std. 239.18
Designed using Perform Pro, WHS/DIOR, Oct 94

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to *stay within the lines* to meet *optical scanning requirements.*

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available
(e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C - Contract	PR - Project
G - Grant	TA - Task
PE - Program	WU - Work Unit
Element	Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with....; Trans. of....; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE - See authorities.
NASA - See Handbook NHB 2200.2.
NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - Leave blank.
DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
Leave blank.
NASA - Leave blank.
NTIS -

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

1998 SREP Final Technical Report Table of Contents

Armstrong Research Laboratory

Volume 1

1998 SREP Final Technical Report Table of Contents

Phillips Research Laboratory
Volume 2

1998 SREP Final Technical Report Table of Contents

Rome Research Laboratory

Volume 3

1998 SREP Final Technical Report Table of Contents

Wright Research Laboratory

Volume 4

Volume 5

1998 SUMMER RESEARCH EXTENSION PROGRAM (SREP) MANAGEMENT REPORT

1.0 BACKGROUND

Under the provisions of Air Force Office of Scientific Research (AFOSR) contract F49620-90-C-0076, September 1990, Research & Development Laboratories (RDL), an 8(a) contractor in Culver City, CA, manages AFOSR's Summer Research Program. This report is issued in partial fulfillment of that contract (CLIN 0003AC).

The Summer Research Extension Program (SREP) is one of four programs AFOSR manages under the Summer Research Program. The Summer Faculty Research Program (SFRP) and the Graduate Student Research Program (GSRP) place college-level research associates in Air Force research laboratories around the United States for 8 to 12 weeks of research with Air Force scientists. The High School Apprenticeship Program (HSAP) is the fourth element of the Summer Research Program, allowing promising mathematics and science students to spend two months of their summer vacations working at Air Force laboratories within commuting distance from their homes.

SFRP associates and exceptional GSRP associates are encouraged, at the end of their summer tours, to write proposals to extend their summer research during the following calendar year at their home institutions. AFOSR provides funds adequate to pay for SREP subcontracts. In addition, AFOSR has traditionally provided further funding, when available, to pay for additional SREP proposals, including those submitted by associates from Historically Black Colleges and Universities (HBCUs) and Minority Institutions (MIs). Finally, laboratories may transfer internal funds to AFOSR to fund additional SREPs. Ultimately the laboratories inform RDL of their SREP choices, RDL gets AFOSR approval, and RDL forwards a subcontract to the institution where the SREP associate is employed. The subcontract (see Appendix 1 for a sample) cites the SREP associate as the principal investigator and requires submission of a report at the end of the subcontract period.

Institutions are encouraged to share costs of the SREP research, and many do so. The most common cost-sharing arrangement is reduction in the overhead, fringes, or administrative charges institutions would normally add on to the principal investigator's or research associate's labor. Some institutions also provide other support (e.g., computer run time, administrative assistance, facilities and equipment or research assistants) at reduced or no cost.

When RDL receives the signed subcontract, we fund the effort initially by providing 90% of the subcontract amount to the institution (normally $18,000 for a $20,000 SREP). When we receive the end-of-research report, we evaluate it administratively and send a copy to the laboratory for a technical evaluation. When the laboratory notifies us the SREP report is acceptable, we release the remaining funds to the institution.

2.0 THE 1998 SREP PROGRAM

SELECTION DATA: A total of 490 faculty members (SFRP Associates) and 202 graduate students (GSRP associates) applied to participate in the 1998 Summer Research Program. From these applicants 188 SFRPs and 98 GSRPs were selected. The education level of those selected was as follows:

1997 SRP Associates, by Degree			
SFRP		GSRP	
PHD	MS	MS	BS
184	6	2	53

Of the participants in the 1997 Summer Research Program 90 percent of SFRPs and 13 percent of GSRPs submitted proposals for the SREP. One undred and thirty-two proposals from SFRPs and seventeen from GSRPs were selected for funding, which equates to a selection rate of 54% of the SFRP proposals and of 34% for GSRP proposals.

1998 SREP: Proposals Submitted vs. Proposals Selected			
	Summer 1997 Participants	Submitted SREP Proposals	SREPs Funded
SFRP	188	132	20
GSRP	98	17	4
TOTAL	286	149	24

The funding was provided as follows:

Contractual slots funded by AFOSR	18
Laboratory funded	22
Total	40

Twelve HBCU/MI associates from the 1997 summer program submitted SREP proposals; six were selected (none were lab-funded; all were funded by additional AFOSR funds).

Proposals Submitted and Selected, by Laboratory		
	Applied	Selected
Armstrong Research Site	9	3
Air Logistic Centers	31	5
Arnold Engineering Development Center	2	1
Phillips Research Site	30	10
Rome Research Site	29	12
Wilford Hall Medical Center	1	0
Wright Research Site	47	9
TOTAL	149	40

Note: Armstrong Research Site funded 1 SREP; Phillips Research Site funded 6; Rome Research Site funded 9; Wright Research Site funded 6.

The 125 1997 Summer Research Program participants represented 60 institutions.

Institutions Represented on the 1997 SRP and 1998 SREP		
Number of schools represented in the Summer 97 Program	Number of schools represented in submitted proposals	Number of schools represented in Funded Proposals
125	110	55

Thirty schools had more than one participant submitting proposals.

Proposals Submitted Per School
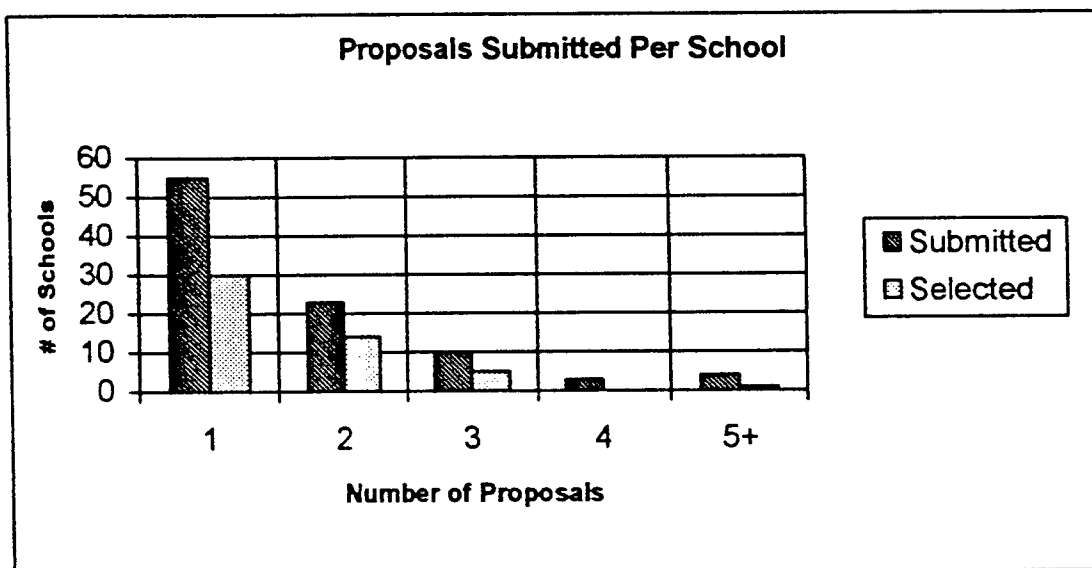
The selection rate for the 65 schools submitting 1 proposal (68%) was better than those submitting 2 proposals (61%), 3 proposals (50%), 4 proposals (0%) or 5+ proposals (25%). The 4 schools that submitted 5+ proposals accounted for 30 (15%) of the 149 proposals submitted.

Of the 149 proposals submitted, 130 offered institution cost sharing. Of the funded proposals which offered cost sharing, the minimum cost share was $3046.00, the maximum was $39,261.00 with an average cost share of $11,069.21.

Proposals and Institution Cost Sharing		
	Proposals Submitted	Proposals Funded
With cost sharing	117	32
Without cost sharing	32	8
Total	149	40

The SREP participants were residents of 31 different states. Number of states represented at each laboratory were:

States Represented, by Proposals Submitted/Selected per Laboratory		
	Proposals Submitted	Proposals Funded
Armstrong Laboratory	31	5
Air Logistic Centers	9	3
Arnold Engineering Development Center	2	1
Phillips Laboratory	30	10
Rome Laboratory	29	12
Wilford Hall Medical Center	1	0
Wright Laboratory	47	9

Nine of the 1997 SREP Principal Investigators also participated in the 1998 SREP.

ADMINISTRATIVE EVALUATION: The administrative quality of the SREP associates' final reports was satisfactory. Most complied with the formatting and other instructions provided to them by RDL. Thirty-seveb final reports have been received and are included in this report. The subcontracts were funded by $992,855.00 of Air Force money. Institution cost sharing totaled $354.215.00.

<u>TECHNICAL EVALUATION</u>: The form used for the technical evaluation is provided as Appendix 2. Thirty-five evaluation reports were received. Participants by laboratory versus evaluations submitted is shown below:

	Participants	Evaluations	Percent
Armstrong Laboratory	5	4	95.2
Air Logistic Centers	3	3	100
Arnold Engineering Development Center	1	1	100
Phillips Laboratory	10	10	100
Rome Laboratory	12	12	100
Wright Laboratory	9	5	91.9
Total	40	35	95.0

Notes:

1: Research on four of the final reports was incomplete as of press time so there aren't any technical evaluations on them to process, yet. Percent complete is based upon 20/21 = 95.2%

2: One technical evaluation was not completed because one of the final reports was incomplete as of press time. Percent complete is based upon 18/18 = 100%

The number of evaluations submitted for the 1998 SREP (95.0%) shows a marked improvement over the 1997 SREP submittals (65%).

<u>PROGRAM EVALUATION</u>: Each laboratory focal point evaluated ten areas (see Appendix 2) with a rating from one (lowest) to five (highest). The distribution of ratings was as follows:
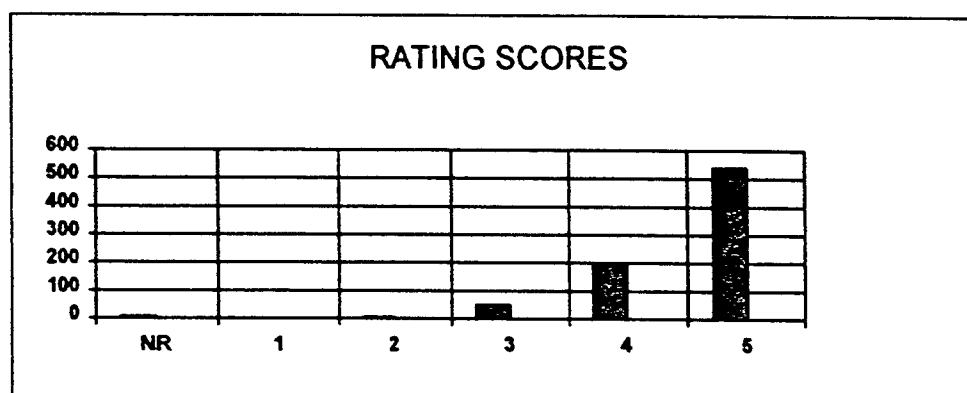
Rating	Not Rated	1	2	3	4	5
# Responses	7	1	7	62 (6%)	226 (25%)	617 (67%)

The 8 low ratings (one 1 and seven 2's) were for question 5 (one 2) "The USAF should continue to pursue the research in this SREP report" and question 10 (one 1 and six 2's) "The one-year period for complete SREP research is about right", in addition over 30% of the threes (20 of 62) were for question ten. The average rating by question was:

Question	1	2	3	4	5	6	7	8	9	10
Average	4.6	4.6	4.7	4.7	4.6	4.7	4.8	4.5	4.6	4.0

The distribution of the averages was:

AREA AVERAGES
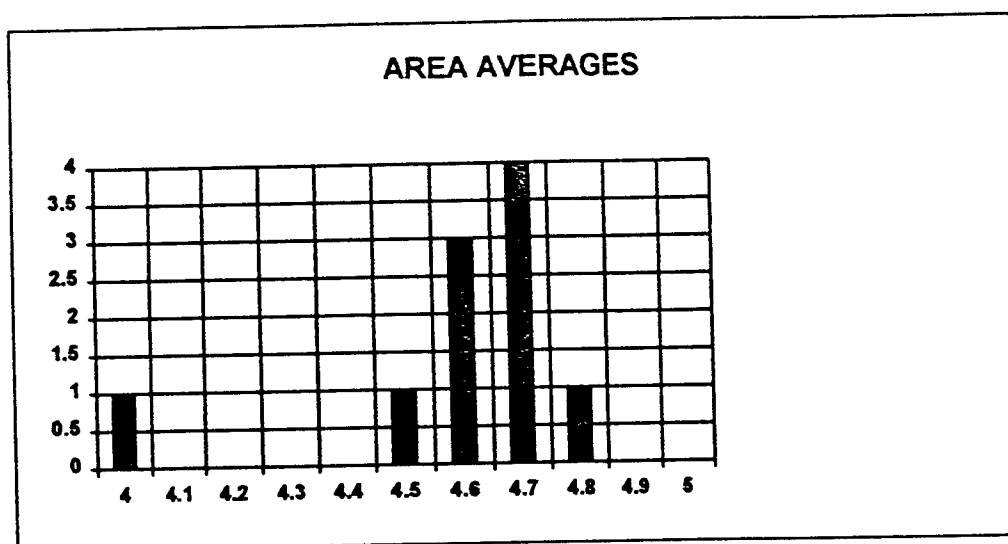
Area 10 "the one-year period for complete SREP research is about right" had the lowest average rating (4.1). The overall average across all factors was 4.6 with a small sample standard deviation of 0.2. The average rating for area 10 (4.1) is approximately three sigma lower than the overall average (4.6) indicating that a significant number of the evaluators feel that a period of other than one year should be available for complete SREP research.

The average ratings ranged from 3.4 to 5.0. The overall average for those reports that were evaluated was 4.6. Since the distribution of the ratings is not a normal distribution the average of 4.6 is misleading. In fact over half of the reports received an average rating of 4.8 or higher. The distribution of the average report ratings is as shown:

AVERAGE RATINGS
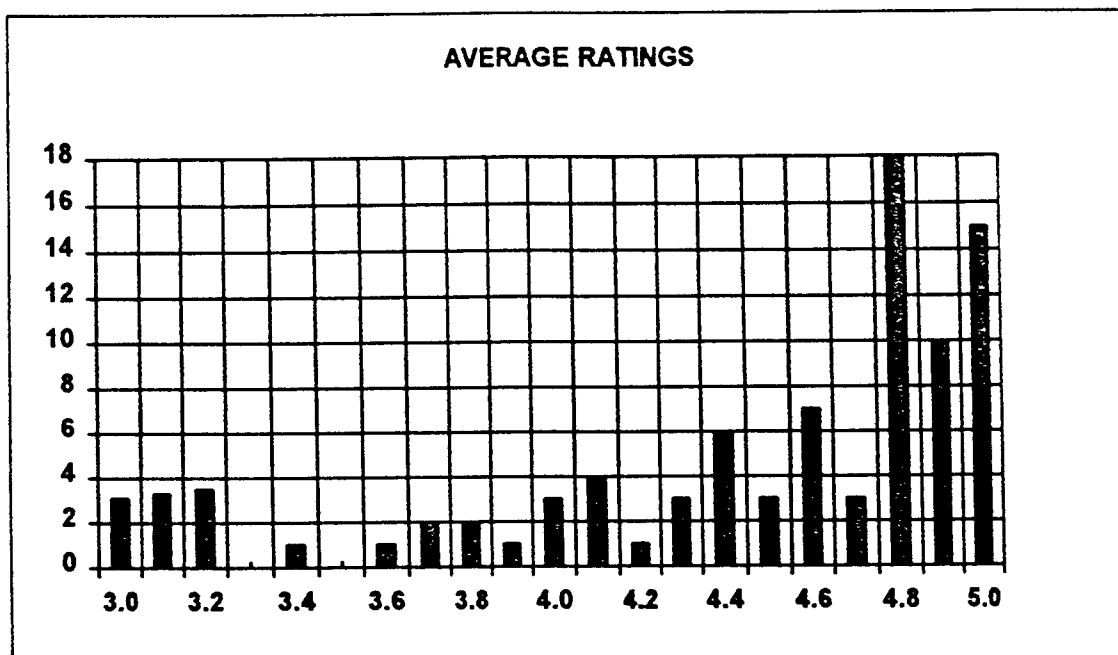
It is clear from the high ratings that the laboratories place a high value on AFOSR's Summer Research Extension Programs.

3.0 SUBCONTRACTS SUMMARY

Table 1 provides a summary of the SREP subcontracts. The individual reports are published in volumes as shown:

Laboratory	Volume
Armstrong Research Site	1
Arnold Engineering Development Center	5
Air Logistic Centers	5
Phillips Research Site	2
Rome Research Site	3
Wright Research Site	4

SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period	Contract Amount	Univ. Cost Share
Chubb , Gerald Industrial Engineering Ohio State University, Columbus, OH	PhD 98-0829	AL/HR	01/01/98 12/31/98 Scoring Pilot Performance of Basic Flight Manuevers	$25000.00	$0.00
Foy , Brent Medical Physics Wright State University, Dayton, OH	PhD 98-0828	AL/OE	01/01/98 12/31/98 Development & Validation of a Physiologically-Based Kinetic Model of Perfused	$25000.00	$11278.00
Lance , Charles Psychology Univ of Georgia Res Foundation, Athens, GA	PhD 98-0842	AL/HR	01/01/98 12/31/98 Extension of Job Performance Measurement Tech to the Development of a Prototype	$24989.00	$0.00
Woehr , David Department of Psychology Texas A & M Univ-College Station, College	PhD 98-0802	AL/HR	01/01/98 12/31/98 Validation of The Multidimensional work ethic profile (MWEP) as a screening too	$25000.00	$11508.00
Collins , Frank Mechanical Engineering Tennessee Univ Space Institute, Tullahoma, TN	PhD 98-0807	AEDC/E	01/01/98 12/31/98 Monte Carlo Computation of SpeciesSeparation by a Conical Skimmer in Hypersonic	$25000.00	$16104.00
Whaley , Paul Mechanical Engineering Oklahoma Christian Univ of Science & Art,	PhD 98-0820	ALC/OC	01/01/98 12/31/98 Probabilistic Analysis of Residual Strength in Corroded and Uncorroded Aging Air	$23351.00	$3046.00
Balas , Mark Applied Math Univ of Colorado at Boulder, Boulder, CO	PhD 98-0816	PL/SX	01/01/98 12/31/98 Non-Linear Adaptive Control for a Precision Deployable Structure with White ligh	$25000.00	$0.00
Duric , Neb Astrophysics University of New Mexica, Albuquerque, NM	PhD 98-0808	PL/LI	01/01/98 12/31/98 Image Recovery Using Phase Diversity	$25000.00	$5777.00
Hanson , George Electrical Engineering Univ of Wisconsin - Milwaukee, Milwaukee, WI	PhD 98-0811	PL/WS	01/01/98 12/31/98 Perturbation Analysis of the Natural Frequencies Targets in Inhomogeneous Media	$25000.00	$23250.00
Jeffs , Brian Electrical Engineering Brigham Young University, Provo, UT	PhD 98-0813	PL/LI	01/01/98 12/31/98 Bayesian Restoration of Space object Images From Adaptive Optics Data with unkno	$25000.00	$19177.00
Kar , Aravinda Engineering University of Central Florida, Orlando, FL	PhD 98-0812	PL/LI	01/01/98 12/31/98 Effects of Vapor-Plasma Layer on Thick-Section Cutting and Calculation of Modes	$25000.00	$5414.00
Leo , Donald Mechanical & Aerospace University of Toledo, Toledo, OH	PhD 98-0810	PL/VT	01/01/98 09/30/98 Adaptive vibration suppression for autonomous Control Systems	$24964.00	$9628.00
Liu , Hanli Physics Univ of Texas at Arlington, Arlington, TX	PhD 98-0814	PL/LI	01/01/98 12/31/98 Continuous-Wave Approach to 3-D Imaging through Turbid media w/a Single Planar M	$25000.00	$11000.00
Bienfang , Joshua Physics University of New Mexico, Albuquerque, NM	BS 98-0815	PL/LI	01/01/98 12/31/98 Optical Clocks Based on Diode Lasers	$24994.00	$0.00
Paulson , Eric Engineering/Physics Univ of Colorado at Boulder, Boulder, CO	BS 98-0837	PL/RK	01/01/98 12/31/98 Optimization & Analysis of a Waverider Vehicle for Global Spaceplane Trajectorie	$25000.00	$7794.00

SREP SUB-CONTRACT DATA

Report Author / Author's University	Author's Degree	Sponsoring Lab	Performance Period	Contract Amount	Univ. Cost Share
Stephens II , Kenneth / University of North Texas, Denton, TX	MA / 98-0809	PL/WS / Simulation of an Explosively Formed Fuse Using MACH 2	01/01/98 12/31/98	$25000.00	$16764.00
Barjaktarovic , Milica / Electrical Engineering / Wilkes University, Wilkes Barre, PA	PhD / 98-0824	RL/IW / Specification and Verification of SDN.701 MSP Functions and Missi Crypto Functio	01/01/98 12/31/98	$24976.00	$3158.00
Batalama , Stella / EE / SUNY Buffalo, Buffalo, NY	PhD / 98-0823	RL/C3 / Robust Spread Spectrum Communications:Adaptive Interference Mitigation Technique	01/01/98 12/31/98	$25000.00	$5600.00
Bourbakis , Nikolaos / Computer Science & Engr / SUNY Binghamton, Binghamton, NY	PhD / 98-0832	RL/IR / hierarchical-Adaptive Image Segmentation	01/01/98 12/31/98	$25000.00	$22723.00
Dasigi , Venugopala / Computer Science / Southern Polytechnic State Univ, Marietta, GA	PhD / 98-0830	RL/C3 / Information Fusion w/Multiple Feature Extractors for automatic Text Classificati	01/01/98 12/31/98	$25000.00	$4000.00
Eckert , Richard / Physics / SUNY Binghamton, Binghamton, NY	PhD / 98-0825	RL/C3 / The Interactive Learning Wall; A PC-Based, Deployable Data Wall for Use in a Co	01/01/98 12/31/98	$25000.00	$39261.00
Lin , Kuo-Chi / Aerospace Engineering / University of Central Florida, Orlando, FL	PhD / 98-0822	RL/IR / Web-Based Distributed Simulation	01/01/98 12/31/98	$25000.00	$0.00
Pados , Dimitrios / Dept. of Electrical /Computer Eng. / State Univ. of New York Buffalo, Buffalo, NY	PhD / 98-0818	RL/OC / Adaptive Array Radars and Joint Space-Time Auxiliary Verctor Filtering	01/01/98 12/31/98	$25000.00	$5600.00
Panda , Brajendra / Computer Science / University of North Dakota, Grand Forks, ND	PhD / 98-0821	RL/CA / Information Warfare: Design of an Efficient Log Management Method to Aid In Dat	01/01/98 12/31/98	$25000.00	$7113.00
Pittarelli , Michael / Systems Science / SUNY OF Tech Utica, Utica, NY	PhD / 98-0827	RL/C3 / Complexity of Detecting and content-driven methods for resolving database incons	01/01/98 12/31/98	$24998.00	$0.00
Schmalz , Mark / Dept of Computer & Info Science / University of Florida, Gainesville, FL	PhD / 98-0831	RL/IR / Errors Inherent in 3D Target Reconstruction from Multiple Airborne Images	01/01/98 12/31/98	$24619.00	$0.00
Ye , Nong / Industrial Engineering / Arizona State University, Tempe, AZ	PhD / 98-0826	RL/CA / Model-Based Assessment of Campaign Plan Performance under Uncertainty	01/01/98 12/31/98	$25000.00	$5000.00
Bradley , Parker / Physics / Syracuse University, Syracuse. NY	BS / 98-0834	RL/IR / Development of User-Friendly Comp Environment for Blind Source Separation Studie	01/01/98 12/31/98	$25000.00	$0.00
Kumar , Devendra / Computer Science / CUNY-City College, New York, NY	PhD / 98-0805	ALC/SA / Further Development of a Simpler, Multiversion concurrency Control Protocol for	01/01/98 12/31/98	$25000.00	$11362.00
Chow , Joe / Mechanical Engineering / Florida International Univ, Miami, FL	PhD / 98-0806	ALC/W / An Automated 3-D Surface Model Creation Module for Laser Scanned Point Data	01/01/98 12/31/98	$25000.00	$5360.00

SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period	Contract Amount	Univ. Cost Share
Beecken , Brian Physics Bethel College. St. Paul, MN	PhD 98-0804	WL/MN	01/01/98 12/31/98	$19986.00	$3997.00
		Development of a statistical Model predicting the impact of a scene projector's			
Beggs , John Electrical Engineering Mississippi State University. Mississippi State,	PhD 98-0817	WL/FI	01/01/98 12/31/98	$25000.00	$25174.00
		Implementation of an Optimization Algorithm in Electromagnentics for Radar Absor			
Bhatnagar , Raj Computer Science University of Cincinnati, Cincinnati, OH	PhD 98-0819	WL/AA	01/01/98 09/30/98	$25000.00	$17488.00
		Analysis of Intra-Class Variability & synthetic Target Models for Use in ATR			
Blaisdell , Gregory Mechanical Engineering Purdue University, West Lafayette, IN	PhD 98-0839	WL/FI	01/01/98 12/31/98	$25000.00	$11844.00
		Validation of a Large Eddy Simulation Code & Development of Commuting Filters			
Douglass , John Zoology University of Arizona, Tucson, AZ	PhD 98-0803	WL/MN	01/01/98 12/31/98	$25000.00	$3719.00
		Roles of Matched Filtering and Coarse in Insect Visual Processing			
Hosford , William Mettalurgy Univ of Michigan, Ann Arbor, MI	PHD 98-0840	WL/MN	01/01/98 12/31/98	$25000.00	$5000.00
		Prediction of Compression Textures in Tantalum Using a Pencil-Glide Computer Mod			
Pan , Yi Computer Science University of Dayton, Dayton, OH	PhD 98-0838	WL/FI	01/01/98 12/31/98	$25000.00	$9486.00
		Parallelization of Time-Dependent Maxwell Equations Using High Perform. Fortran			
Pochiraju , Kishore Mechanical Engineering Stevens Inst of Technology, Hoboken, NJ	PhD 98-0833	WL/ML	01/01/98 12/31/98	$25000.00	$9625.00
		A Hybrid Variational-Asymptotic Method for the Analysis of MicroMechanical Damag			
Shtessel , Yuri Electrical Engineering Univ of Alabama at Huntsville, Huntsville, AL	PhD 98-0841	WL/FI	01/01/98 12/31/98	$25000.00	$4969.00
		Continuous Sliding Mode Control Approach for Addressing actutor Deflection and			
Starzyk , Janusz Electrical Engineering Ohio University, Athens, OH	PhD 98-0801	WL/AA	01/01/98 12/31/98	$24978.00	$12996.00
		Feature Selection for Automatic Target Recognition:Mutual Info & Stat Tech			

APPENDIX 1:

SAMPLE SREP SUBCONTRACT

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
1998 SUMMER RESEARCH EXTENSION PROGRAM
SUBCONTRACT 98-0812

BETWEEN

Research & Development Laboratories
5800 Uplander Way
Culver City, CA 90230-6608

AND

University of Central Florida
Office of Sponsored Research/ Admin#423
4000 Central Florida Blvd.
Orlando, FL 32816-0150

REFERENCE: Summer Research Extension Program Proposal 97-0018

Start Date: 01/01/98 End Date: 12/31/98

Proposal Amount: $25000.0
Proposal Title:
Effects of Vapor-Plasma Layer on Thick-Section Cutting and Calculation of Modes

(1) PRINCIPAL INVESTIGATOR:

DR Aravinda Kar
CREOL
University of Central Florida
Orlando, FL 32816-2700

(2) UNITED STATES AFOSR CONTRACT NUMBER: F49620-93-C-0063

(3) CATALOG OF FEDERAL DOMESTIC ASSISTANCE NUMBER (CFDA): 12.800
PROJECT TITLE: AIR FORCE DEFENCE RESEARCH SOURCES PROGRAM

(4) ATTACHMENTS
1 REPORT OF INVENTIONS AND SUBCONTRACT
2 CONTRACT CLAUSES
3 FINAL REPORT INSTRUCTIONS

*** SIGN SREP SUBCONTRACT AND RETURN TO RDL ***

1. BACKGROUND: Research & Development Laboratories (RDL) is under contract (F49620-93-C-0063) to the United States Air Force to administer the Summer Research Program (SRP), sponsored by the Air Force Office of Scientific Research (AFOSR), Bolling Air Force Base, D.C. Under the SRP, a selected number of college faculty members and graduate students spend part of the summer conducting research in Air Force laboratories. After completion of the summer tour participants may submit, through their home institutions, proposals for follow-on research. The follow-on research is known as the Summer Research Extension Program (SREP). Approximately 61 SREP proposals annually will be selected by the Air Force for funding of up to $25,000; shared funding by the academic institution is encouraged. SREP efforts selected for funding are administered by RDL through subcontracts with the institutions. This subcontract represents an agreement between RDL and the institution herein designated in Section 5 below.

2. RDL PAYMENTS: RDL will provide the following payments to SREP institutions:

   - 80 percent of the negotiated SREP dollar amount at the start of the SREP research period.

   - The remainder of the funds within 30 days after receipt at RDL of the acceptable written final report for the SREP research.

3. INSTITUTION'S RESPONSIBILITIES: As a subcontractor to RDL, the institution designated on the title page will:

a. Assure that the research performed and the resources utilized adhere to those defined in the SREP proposal.

b. Provide the level and amounts of institutional support specified in the SREP proposal..

c. Notify RDL as soon as possible, but not later than 30 days, of any changes in 3a or 3b above, or any change to the assignment or amount of participation of the Principal Investigator designated on the title page.

d. Assure that the research is completed and the final report is delivered to RDL not later than twelve months from the effective date of this subcontract, but no later than December 31, 1998. The effective date of the subcontract is one week after the date that the institution's contracting representative signs this subcontract, but no later than January 15, 1998.

e. Assure that the final report is submitted in accordance with Attachment 3.

f. Agree that any release of information relating to this subcontract (news releases, articles, manuscripts, brochures, advertisements, still and motion pictures, speeches, trade associations meetings, symposia, etc.) will include a statement that the project or effort depicted was or is sponsored by: Air Force Office of Scientific Research, Bolling AFB, D.C.

g. Notify RDL of inventions or patents claimed as the result of this research as specified in Attachment 1.

h. RDL is required by the prime contract to flow down patent rights and technical data requirements to this subcontract. Attachment 2 to this subcontract

contains a list of contract clauses incorporated by reference in the prime contract.

4. All notices to RDL shall be addressed to:

> RDL AFOSR Program Office
> 5800 Uplander Way
> Culver City, CA 90230-6609

5. By their signatures below, the parties agree to provisions of this subcontract.

_____
Abe Sopher
RDL Contracts Manager

_____
Date

_____
Signature of Institution Contracting Official

_____
Typed/Printed Name

_____
Title

_____
Institution

_____
Date Phone

ATTACHMENT 2
CONTRACT CLAUSES

This contract incorporates by reference the following clauses of the Federal Acquisition Regulations (FAR), with the same force and effect as if they were given in full text. Upon request, the Contracting Officer or RDL will make their full text available (FAR 52.252-2).

FAR CLAUSES	TITLE AND DATE
52.202-1	DEFINITIONS
52.203-3	GRATUITIES
52.203-5	COVENANT AGAINST CONTINGENT FEES
52.203-6	RESTRICTIONS ON SUBCONTRACTOR SALES TO THE GOVERNMENT
52.203-7	ANTI-KICKBACK PROCEDURES
52.203-8	CANCELLATION, RECISSION, AND RECOVERY OF FUNDS FOR ILLEGAL OR IMPROPER ACTIVITY
52.203-10	PRICE OR FEE ADJUSTMENT FOR ILLEGAL OR IMPROPER ACTIVITY
52.203-12	LIMITATION ON PAYMENTS TO INFLUENCE CERTAIN FEDERAL TRANSACTIONS
52.204-2	SECURITY REQUIREMENTS
52.209-6	PROTECTING THE GOVERNMENT'S INTEREST WHEN SUBCONTRACTING WITH CONTRACTORS DEBARRED, SUSPENDED, OR PROPOSED FOR DEBARMENT
52.212-8	DEFENSE PRIORITY AND ALLOCATION REQUIREMENTS
52.215-2	AUDIT AND RECORDS - NEGOTIATION
52.215-10	PRICE REDUCTION FOR DEFECTIVE COST OR PRICING DATA

52.215-12	SUBCONTRACTOR COST OR PRICING DATA
52.215-14	INTEGRITY OF UNIT PRICES
52.215-8	ORDER OF PRECEDENCE
52.215.18	REVERSION OR ADJUSTMENT OF PLANS FOR POSTRETIREMENT BENEFITS OTHER THAN PENSIONS
52.222-3	CONVICT LABOR
52.222-26	EQUAL OPPORTUNITY
52.222-35	AFFIRMATIVE ACTION FOR SPECIAL DISABLED AND VIETNAM ERA VETERANS
52.222-36	AFFIRMATIVE ACTION FOR HANDICAPPED WORKERS
52.222-37	EMPLOYMENT REPORTS ON SPECIAL DISABLED VETERAN AND VETERANS OF THE VIETNAM ERA
52.223-2	CLEAN AIR AND WATER
52.223-6	DRUG-FREE WORKPLACE
52.224-1	PRIVACY ACT NOTIFICATION
52.224-2	PRIVACY ACT
52.225-13	RESTRICTIONS ON CONTRACTING WITH SANCTIONED PERSONS
52.227-1	ALT. I - AUTHORIZATION AND CONSENT
52.227-2	NOTICE AND ASSISTANCE REGARDING PATIENT AND COPYRIGHT INFRINGEMENT

DOD FAR CLAUSES	DESCRIPTION
252.203-7001	SPECIAL PROHIBITION ON EMPLOYMENT
252.215-7000	PRICING ADJUSTMENTS
252.233-7004	DRUG FREE WORKPLACE (APPLIES TO SUBCONTRACTS WHERE THERE IS ACCESS TO CLASSIFIED INFORMATION)
252.225-7001	BUY AMERICAN ACT AND BALANCE OF PAYMENTS PROGRAM
252.225-7002	QUALIFYING COUNTRY SOURCES AS SUBCONTRACTS
252.227-7013	RIGHTS IN TECHNICAL DATA - NONCOMMERCIAL ITEMS
252.227-7030	TECHNICAL DATA - WITHOLDING PAYMENT
252.227-7037	VALIDATION OF RESTRICTIVE MARKINGS ON TECHNICAL DATA
252.231-7000	SUPPLEMENTAL COST PRINCIPLES
252.232-7006	REDUCTIONS OR SUSPENSION OF CONTRACT PAYMENTS UPON FINDING OF FRAUD

APPENDIX 2:

SAMPLE TECHNICAL EVALUATION FORM

SUMMER RESEARCH EXTENSION PROGRAM
TECHNICAL EVALUATION

SREP No: 98-0810
Principal Investigator:DR Donald Leo
University of Toledo

Circle the rating level number, 1 (low) through 5 (high),
you feel best evaluate each statement and return the
completed form to RDL by fax of mail to:

RDL
Attn: SREP Tech Evals
5800 Uplander Way
Culver City, CA 90230-6608
(310)216-5940 or (800)677-1363

1. This SREP report has a high level of technical merit.	1 2 3 4 5
2. The SREP program is important to accomplishing the lab's mission.	1 2 3 4 5
3. This SREP report accomplished what the associate's proposal promised.	1 2 3 4 5
4. This SREP report addresses area(s) important to the USAF.	1 2 3 4 5
5. The USAF should continue to pursue the research in this SREP report.	1 2 3 4 5
6. The USAF should maintain research relationships with this SREP associate.	1 2 3 4 5
7. The money spent on this SREP effort was well worth it.	1 2 3 4 5
8. This SREP report is well organized and well written.	1 2 3 4 5
9. I'll be eager to be a focal point for summer and SREP associates in the future.	1 2 3 4 5
10. The one-year period for complete SREP research is about right.	1 2 3 4 5

11. If you could change any one thing about the SREP program, what would you change:

_____

_____

_____

_____

12. What do you definitely NOT change about the SREP program?

_____

_____

_____

_____

PLEASE USE THE BACK FOR ANY OTHER COMMENTS

Laboratory Phillips Laboratory
Lab Focal Point Capt Jeanne Sullivan
Office Symbol AFRL/VSDV   Phone: (505)846-2069

Introduction - 22

A STATISTICAL MODEL PREDICTING THE IMPACT OF A SCENE PROJECTOR'S NONUNIFORMITY ON A TEST ARTICLE'S IMAGE

Brian P. Beecken
Professor
Department of Physics

Bethel College
3900 Bethel Drive
St. Paul, MN 55112

A STATISTICAL MODEL PREDICTING THE IMPACT OF A SCENE PROJECTOR'S NONUNIFORMITY ON A TEST ARTICLE'S IMAGE

Brian P. Beecken
Professor
Department of Physics
Bethel College

ABSTRACT

A statistical model has been developed and used to calculate the contribution of a scene projector array to the nonuniformity of a test article's output image. The output image nonuniformity is different from the detector array's nonuniformity because of the nonuniformity of the projector array and the relative positions of emitter images on the detector array as determined by the sampling ratio. These two factors are combined with the nonuniformity of the detector array using standard error propagation techniques. In order to calculate the predicted output nonuniformity, a weighting function was developed to determine the contribution of the different emitters upon individual detector elements. It is through this weighting function that parameters such as the sampling ratio, the fill factor of the detector array, the optical blur of the emitters, and the alignment of the emitters with respect to the detectors affect the nonuniformity. A computer program has been written to numerically approximate the weighting function for a user defined set of parameters. The model presented in this paper represents a significant improvement over one reported earlier.[1] The new version includes noninteger sampling ratios and contains a number of provisions for possible experimental concerns. Such issues include detector offset nonuniformity, diffraction, uncertainty in projector nonuniformity, and spatial droop.

A STATISTICAL MODEL PREDICTING THE IMPACT OF A SCENE PROJECTOR'S NONUNIFORMITY ON A TEST ARTICLE'S IMAGE

Brian P. Beecken

Introduction

The KHILS facility at Eglin AFB tests infrared imaging sensors in the lab by projecting simulated IR scenes that subtend the sensor's field of view.[2] The IR projector, an integral component of the simulation testing, is produced under the WISP (Wideband Infrared Scene Projector) program.

The WISP projector consists of an array of emitters at least as large in number as the array of detectors on the sensor's focal plane.[3] Unfortunately, the fact that the projector consists of an array of emitter pixels will cause some nonuniformity in the test article's output image. This circumstance occurs for two reasons. First, when the emitter array is set to produce a uniform IR scene, there is significant nonuniformity in the array's output.[4] Although this nonuniformity can be corrected to a large extent, it can never be completely eliminated. Second, unless there is precisely the same number of emitters for every detector element, some detectors will "see" more emitters (or at least larger portions of them) than other detectors. Thus, the pixelized nature of the projector will exist as an artifact in the test article's output image. This artifact is solely the result of the simulation, and represents a degradation of the actual scene that would be encountered during the sensor's mission.

The goal of this report is to predict analytically how the nonuniformity of the projector impacts the nonuniformity of the test article's output image. A statistical model for this prediction has been developed previously.[1] The present work represents a significant extension of that model so that it includes non-integer sampling ratios. Also considered are detector offset non-uniformity, the difference between a Gaussian approximation and Fraunhofer diffraction, the impact of inaccurate knowledge of the projector nonuniformity, and the effect of spatial droop on the model. Such information will be valuable for determining both the realism of the simulation and the nonuniformity correction requirements. These topics also coincide with those discussed in the proposal for the grant which funded this work.[5] In order to facilitate the use of this report, each of these topics has its own subsection. The exception is non-integer sampling

1-3

ratios, which are such a central part of the statistical model that they can not be separated out.

In order to make our model as general as possible, a number of parameters will be required. These parameters are the projector's nonuniformity, the detector array's nonuniformity, the linear ratio of the emitters to detectors (i.e., the sampling ratio), the relative size of the optical blur on the detector array, the alignment of the emitters relative to the detectors, and the fill factor of the detector array. These are believed to be the test parameters of primary importance and will be incorporated into the following analysis.

1 Theory

1.1 The Problem of Positional Nonuniformity with No Blur

The *sampling ratio n:m* is defined as the ratio of emitters n to detectors m when counted in one dimension. Non-integer sampling ratios cause a type of nonuniformity that we will call *positional nonuniformity*. This type of nonuniformity will increase the nonuniformity of the test article's output image and can occur any time there is a non-integer sampling ratio. It has nothing to do with the emitter array's nonuniformity. It is due solely to the fact that if the sampling ratio is non-integer, then in the vast majority of cases, there will not be an equal number of emitter images over each detector element.

In order to illustrate this issue, consider a non-integer sampling ratio of 4:3. Because we are considering the case of no blur, it is appropriate to represent the image of the point emitters as small, solid dots as shown in Fig. 1. The detector elements are represented as open squares.
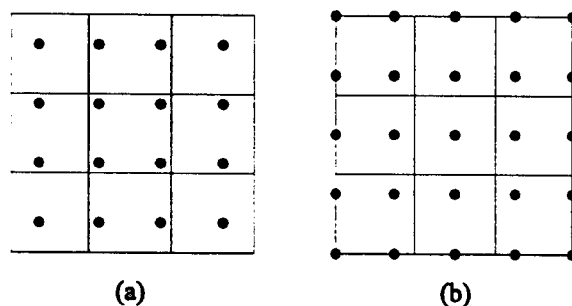

(a) (b)

Figure 1: Illustrations of a sampling ratio of 4:3 with essentially "no blur."

Clearly, in Fig. 1a one detector element has four times as much incident flux as some of the other

1-4

detectors. This arrangement leads to an inherent nonuniformity even if each emitter is producing the same amount of radiation. A translation of half an emitter spacing (diagonally) produces a different pattern as shown in Fig. 1b, an alignment that despite having less nonuniformity still falls far short of being uniform.

Careful consideration of these and many other sampling ratios and alignments leads to the almost obvious conclusion that the patterns will repeat in each direction after m detectors. Thus, the positional nonuniformity pattern will always have a minimum pattern size that is $m \times m$. We will refer to this square as a *chunk*, one two-dimensional spatial period of the nonuniformity pattern across the detector array. A calculation of the positional nonuniformity for this chunk gives the positional nonuniformity for the entire FPA. An algorithm was developed and a computer routine written to rapidly calculate the positional nonuniformity for a large number of non-integer sampling ratios. This calculation was performed for the case where the emitters are spaced as far from the edge of each detector element as possible, such as in Fig. 1a. Many interesting features are apparent in the plot of this calculation shown in Fig. 2. Obviously, the nonuniformity goes to zero when the sampling ratio is an integer. In such cases there is no positional nonuniformity because each detector has the same number of emitters positioned above it. Second, the nonuniformity reaches local maxima between each integer sampling ratio. Third, the maximum positional nonuniformity occurs nearly halfway between the nearest integer ratios.[1] Finally, the effects of positional nonuniformity decrease as the sampling ratio increases.

It turns out that in this kind of analysis alignment plays a very important role. If an emitter is aligned with each corner of the chunk, as for example in Fig. 1b, then rather than a smooth trend, slight changes in sampling ratio cause large and seemingly unpredictable discrete jumps in the nonuniformity (see Fig. 3). In general, those sampling ratios which are the simplest fractions have the most significant drop in nonuniformity with a change in alignment. The two plots shown here represent those alignment cases that have the most positional nonuniformity and the least.

1.2   The General Case

We will now calculate how the test article's nonuniformity depends on detector array nonuniformity, emitter array nonuniformity, and positional nonuniformity. The calculation will lead to

---

[1]As the sampling ratio is increased, each successive maximum approaches more closely to the halfway point between the nearest integer sampling ratios.
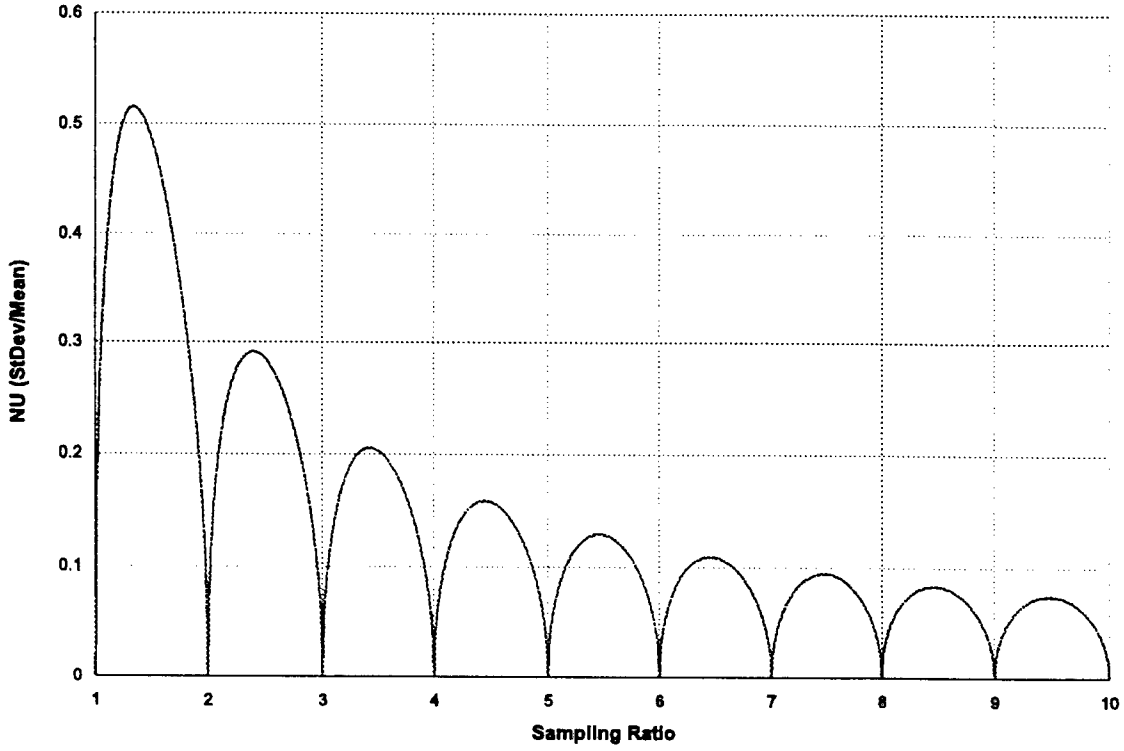
Figure 2: Plot of maximum positional nonuniformity with essentially "no blur."

a general equation that is valid regardless of blur size, alignment, or other such parameters.

The output signal v of a detector is often approximated by

$$v = g\eta\Phi + b \tag{1}$$

where g is the detector's conversion gain, η is the detector's quantum efficiency, Φ is the average number of incident photons per integration time, and b is the signal offset that exists even for zero incident photon flux. The output signal v can only be represented by a linear function over a limited region. It is this region, however, for which two point calibration (offset and gain correction) is performed. In this paper, we will only be considering those situations in which nonuniformity noise dominates over the other noise sources.

Each detector on the array produces a signal that differs from the other detectors if either the conversion gain or the quantum efficiency of the detector varies, or if the incident photon flux is different. We are not interested in whether a variation in signal is due to a detector's conversion
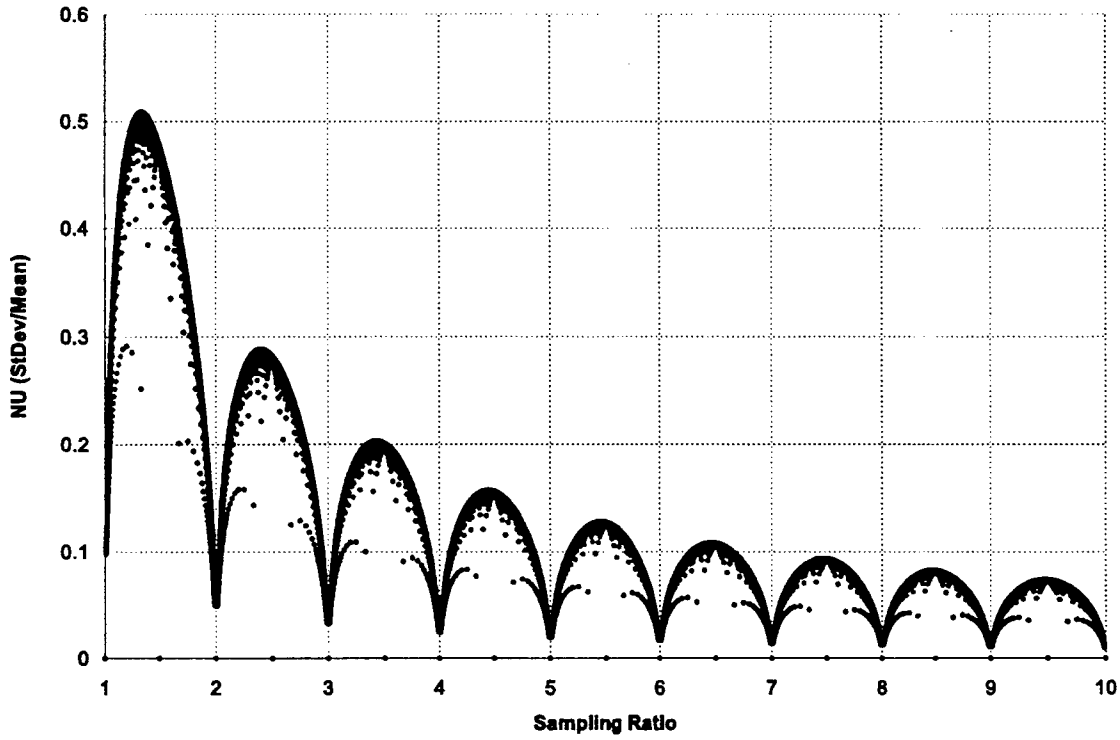
Figure 3: Plot of minimum positional nonuniformity with essentially "no blur."

gain or quantum efficiency. Therefore, these terms will be combined together to represent the detector's sensitivity $g\eta$. The nonuniformity of the detector array can then be conveniently represented as the standard deviation of all the detector sensitivities divided by the average sensitivity, $\frac{\sigma_{g\eta}}{\overline{g\eta}}$. Similarly, the nonuniformity of the emitter array can be represented by $\frac{\sigma_\phi}{\overline{\phi}}$, where $\overline{\phi}$ is the average emittance across the emitter array and σ_ϕ is the standard deviation.

Equation (1) is a product of two terms (detector sensitivity and incident flux) that each have an uncertainty. The standard error propagation equation in this situation is

$$\frac{\sigma_v}{\overline{v}} = \sqrt{\left(\frac{\sigma_{g\eta}}{\overline{g\eta}}\right)^2 + \left(\frac{\sigma_\Phi}{\overline{\Phi}}\right)^2}. \tag{2}$$

So far everything has been straightforward; however, the difficulty at this point is calculating the ratio $\frac{\sigma_\Phi}{\overline{\Phi}}$, which represents the nonuniformity of the radiation incident on the detector array. Obviously this will depend to some extent on the emitter array nonuniformity $\frac{\sigma_\phi}{\overline{\phi}}$, but it will also depend on the positional nonuniformity discussed above in Subsection 1.1. First, let us

1-7

carefully define some terms:

$$\Phi_{ij} \equiv \text{number of photons incident on detector } i, j \text{ during the integration time.}$$

$$\phi_{\gamma\delta} \equiv \text{number of photons emitted by emitter } \gamma, \delta \text{ during the integration time.}$$

$$\omega_{\gamma\delta} \equiv \text{weighting function denoting the fraction of the total flux from}$$

$$\text{a particular emitter } \gamma, \delta \text{ which lands on a particular detector } i, j.$$

$$\text{(3)}$$

With these definitions, the total flux incident on a particular detector is given by

$$\Phi_{ij} = \sum_{\gamma,\delta} \omega_{\gamma\delta} \phi_{\gamma\delta}. \tag{4}$$

Using standard error propagation technique, we have

$$\sigma_{\Phi_{ij}}^2 = \sum_{\gamma,\delta} \left[\left(\frac{\partial \Phi_{ij}}{\partial \omega_{\gamma\delta}} \right)^2 \sigma_{\omega_{\gamma\delta}}^2 + \left(\frac{\partial \Phi_{ij}}{\partial \phi_{\gamma\delta}} \right)^2 \sigma_{\phi_{\gamma\delta}}^2 \right] \tag{5}$$

which reduces to

$$\sigma_{\Phi_{ij}}^2 = \sum_{\gamma,\delta} \left[(\phi_{\gamma\delta})^2 \sigma_{\omega_{\gamma\delta}}^2 + (\omega_{\gamma\delta})^2 \sigma_{\phi_{\gamma\delta}}^2 \right]. \tag{6}$$

Each individual emitter's flux $\phi_{\gamma\delta}$ is approximated by the average flux $\overline{\phi}$ of all the emitters. Similarly, all the uncertainties in the emitted flux $\sigma_{\phi_{\gamma\delta}}$ are the same, because they are all represented by the same standard deviation of the emittances across the emitter array. Thus, $\sigma_{\phi_{\gamma\delta}} = \sigma_{\phi}$. Furthermore, because the weighting function $\omega_{\gamma\delta}$ will be somewhat different for every detector i, j in the chunk, the sum of the weighting functions squared will be slightly different. Therefore, we must take the average of the summation that appears in the second term. Then the calculation will give the appropriate σ_{Φ}^2 for all detectors in the chunk:

$$\sigma_{\Phi}^2 = \overline{\phi}^2 \sum_{\gamma,\delta} \sigma_{\omega_{\gamma\delta}}^2 + \sigma_{\phi}^2 \overline{\sum_{\gamma,\delta} \omega_{\gamma\delta}^2}. \tag{7}$$

It is well-known that the sum of the variances is equal to the variance of the sum. Thus,

$$\sigma_{\Phi}^2 = \overline{\phi}^2 \sigma_{\sum \omega_{\gamma\delta}}^2 + \sigma_{\phi}^2 \overline{\sum_{\gamma,\delta} \omega_{\gamma\delta}^2}. \tag{8}$$

In this equation, the first term represents the positional nonuniformity and the second is the emitter nonuniformity. Realizing that the value for the average flux incident on a detector should be

$$\overline{\Phi} = \overline{\phi} \, \overline{\sum_{\gamma,\delta} \omega_{\gamma\delta}}, \tag{9}$$

we can substitute Eqs. (8) and (9) into Eq. (2),

$$\frac{\sigma_v}{\overline{v}} = \sqrt{\underbrace{\left(\frac{\sigma_{g\eta}}{\overline{g\eta}}\right)^2}_{\text{Det NU}} + \underbrace{\left(\frac{\sigma_{\sum \omega_{\gamma\delta}}}{\overline{\sum \omega_{\gamma,\delta}}}\right)^2}_{\text{Pos NU}} + \underbrace{\left(\frac{\sigma_\phi}{\overline{\phi}}\right)^2}_{\text{Emit NU}} \underbrace{\frac{\overline{\sum \omega_{\gamma\delta}^2}}{\overline{\sum \omega_{\gamma\delta}}^2}}_{\text{Weighting}}}. \qquad (10)$$

This result is very satisfying because the three contributions to the output image nonuniformity—detector array nonuniformity, emitter array nonuniformity, and the nonuniformity arising from the relative positions of the emitters to the detectors—are all easily identifiable. With the exception of the weighting factor, the noise terms simply add. The weighting function is critical to using this equation. The determination of it will be discussed in Section 2.

1.3 Detector Offset Nonuniformity

One of the concerns that has been raised about our approach is the use of Eq. (1). It has been correctly pointed out that each detector element will not only have a different conversion gain g and quantum efficiency η, but will also have a different signal offset b. Thus, at first glance it would seem that we are only accounting for two of the three sources of performance differences between detector elements.

Such a concern is not relevant if all detector elements have undergone a two-point calibration and are being operated within the range of the calibration. Two-point calibration sets all detectors to give the same output signal at two points, one at a low incident flux and the other at a high incident flux. The first point is the offset. The amplifier gain is then set so that all detectors have the same signal response at the second point. Often this calibration is referred to as "offset and gain correction." Thus, after calibration each detector should have the same value of b. Why then is there nonuniformity in $g\eta$? Each detector will respond to increased incident flux in different ways. Only when the increased flux is the same as the second point of the calibration do all the detectors have the same response. If all the detectors had a linear response, then the calibration would eliminate all detector nonuniformity. Since the detectors' behavior is nonlinear, the detector array only exhibits uniformity at the two points at which the calibration was performed. This situation is illustrated in Fig. 4.

The use of $\sigma_{g\eta}$ to represent the uncertainty of a particular detector's response is an appropriate representation of the various responses possible when the incident flux levels are between the

1-9
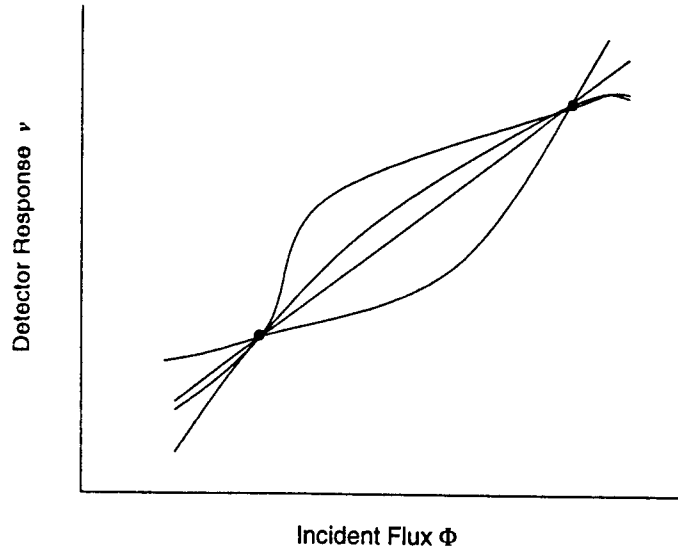
Figure 4: Illustration of two point calibration for four different detectors. Only one of the detectors is linear. The different detector responses between the two calibration points is quantified by $\sigma_{g\eta}$.

calibration points. Outside of this region, neither this paper's calculation of expected nonuniformity nor the calibration itself is valid. If, for instance, someone were to measure the array's nonuniformity below the offset, significant and unpredictable nonuniformity would result.

2 Procedure for Determining the Weighting Function

In the previous section, we derived Eq. (10), which predicts the nonuniformity of the detector's output image. In order to use this equation, however, it is necessary to know the weighting function $\omega_{\gamma\delta}$ as defined in Definition (3). Unfortunately, the weighting function cannot be solved analytically. In this section, we will describe our numerical approach to finding the weighting function. We chose to approximate the distribution of the flux incident on the FPA from a single emitter with a two-dimensional Gaussian function. The validity of this approximation will be discussed in the second part of this section.

2.1 Numerical Calculation of the Weighting Function

The portion of the emitter's flux that lands on a detector is determined by integrating the two-dimensional Gaussian over the detector's area. The fraction of each individual emitter's total flux that is incident on a particular detector i, j is what we have defined as $\omega_{\gamma\delta}$, the weighting function.[2] The average number of emitters per detector is simply the square of the sampling ratio, $\left(\frac{n}{m}\right)^2$. Therefore, on average, as the fill factor of the detector array goes to 100%, the sum of the weighting function must approach this value:

$$\overline{\sum_{\gamma, \delta} \omega_{\gamma\delta}} \longrightarrow \left(\frac{n}{m}\right)^2. \tag{11}$$

In this equation, the sum is over all emitters near enough to a particular detector to contribute incident flux. The average is taken over all detectors in the chunk.

Although the idea behind the calculation of the weighting function is relatively simple, it becomes difficult in practice for three reasons. First, since we cannot analytically solve the integral of the two dimensional Gaussian over square areas, we approximated it numerically by doing Riemann sums. Second, the equation for $\omega_{\gamma\delta}$ becomes rather complex once parameters such as fill factor and alignment are taken into account. Third, the weighting function will be different for every detector i, j within the chunk because the positioning of the emitters with respect to the detector is different for each detector. Our approximation for the weighting function is:

$$\omega_{\gamma\delta} \approx \frac{F_x F_y}{2\pi(.2616KC)^2} \sum_{\alpha=1}^{C} \sum_{\beta=1}^{C} e^{-\frac{1}{2}\left\{\left[\frac{\frac{m}{n}(\gamma+T_x)-i+\frac{1}{2}+\frac{F_x}{2}-\frac{F_x}{C}\left(\alpha-\frac{1}{2}\right)}{.2616K}\right]^2 \left[\frac{\frac{m}{n}(\delta+T_y)-j+\frac{1}{2}+\frac{F_y}{2}-\frac{F_y}{C}\left(\beta-\frac{1}{2}\right)}{.2616K}\right]^2\right\}}.$$

$$\tag{12}$$

The integration of the two-dimensional Gaussian is approximated by using a number of "detection points" distributed across the active detector area. The total number of detection points per detector is C^2. The value of the Gaussian at each detection point is multiplied by the fraction $\frac{F_x F_y}{C^2}$ of the detector's area being represented by the detection point (F_x and F_y are discussed below). The sum of these values yields a good approximation of the integral, assuming that a large enough value of C has been chosen. The computer program we developed to calculate $\omega_{\gamma\delta}$ will automatically choose an appropriate value for C when given values for the

---

[2]This definition of the weighting function represents a significant departure from the definition employed in our previous work, Ref. 1.

sampling ratio $\frac{n}{m}$ and the relative optical blur size K (discussed below). Clearly, a larger C will provide a better approximation, but computer processing time increases rapidly with C.

The *relative optical blur size K* is the size of the optical blur relative to the detector elements on the FPA. We have defined K to be the number of detector pitches in the diameter of a circular blur containing 83.9% of the radiation. It has previously been shown[1] that K can be found experimentally using

$$K = 2.44 \frac{\lambda f}{Dw} \tag{13}$$

where f is the focal length, λ is the wavelength of the incident radiation, w is the detector pitch, and D is the diameter of the system's aperture. It has also been shown in Ref. 1 that the σ of the Gaussian can be written as $0.2616K$. This form appears in the weighting function Eq. (12) in place of the usual σ in the two-dimensional Gaussian, both in the exponent and in the normalization factor.

In Eq. (12), T_x and T_y account for the alignment of the emitter array relative to the FPA. Each variable represents a translation in the x or y direction measured as a fraction of the emitter spacing. Zero translation is defined as one emitter being located precisely above each corner of the chunk of detectors.[3] If the values for translation range from 0 to 0.5, all possible patterns will be covered.[4] The detector's fill factor is accounted for by F_x and F_y, which are the linear fill factors in the x and y directions. Each must be in the range of 0 to 1, where 1 is a 100% linear fill factor.

2.2   Gaussian Distribution versus Fraunhofer Diffraction

When light passes through the aperture of an optical system, it undergoes Fraunhofer diffraction. Clearly, a two-dimensional Gaussian distribution is only a rough approximation to the central maximum and the surrounding fringes. The existence of the central bright spot, or Airy disk, is acknowledged widely enough that it seemed to be the most appropriate measurement for blur size. Consequently, in the preceding subsection we defined K in terms of the diameter of a blur containing 83.9% of the incident radiation, the same percentage that exists in the Airy disk.[6] Unfortunately, there is no such thing as an Airy disk for the Gaussian that we have used, and so this criterion is really rather arbitrary.

---

[3]This definition of zero translation represents a significant departure from the definition employed in our previous work, Ref. 1.

[4]In many cases the pattern will repeat with a translation smaller than 0.5.

In response to some concerns expressed at AeroSense 1998, and in an effort to obtain the most accurate calculation of nonuniformity possible, a significant amount of effort was devoted to studying Fraunhofer diffraction in the present context. Fraunhofer diffraction is normally expressed mathematically in terms of Bessel functions. The Gaussian form of the weighting function was replaced with an appropriate expression using Bessel functions. The weighting function could then be easily used to determine what fraction of the radiation from a single emitter falls on all the nearby detectors. This calculation is equivalent to turning on only one emitter and looking at its blur with a detector array.

Interestingly enough, the predicted numbers for the nearby detector elements were such that it was very difficult, but not impossible, to discern a diffraction pattern. As would be expected, the pattern became somewhat more obvious as the resolution was improved by enlarging the blur so that many detector elements were involved. The circular nature of the diffraction pattern was at least partially obscured by the use of square detectors. In addition, it is important to keep in mind that because 83.9% of the radiation is in the central spot, relatively little radiation is left to define the diffraction rings. The implication of these calculations is that it would be very difficult to experimentally observe a diffraction pattern unless the number of detector elements per emitter were unrealistically large.

The form of the weighting function using Bessel functions was incorporated into the computer model. For large blur sizes ($K > 4$) there were no discernable differences from the usual model. The results for small blur sizes, however, were disturbingly inconsistent. The problem increased as the blur size decreased. Eventually, the difficulty was traced to the fact that Bessel functions are only a valid representation of Fraunhofer diffraction in the small angle approximation. A smaller blur size is the equivalent of putting the detector array closer to the aperture. In that configuration, the angles to the detection points become too large to meet the criteria of a small angle approximation. Thus, except for large blur, the Bessel functions are not a valid way of predicting the distribution of the radiation on the detector array. Because this is a problem for blur sizes of practical interest ($0.5 \leq K \leq 2$), it was decided that the Gaussian approach to the weighting function was the better approximation.

One more point about Fraunhofer diffraction must be made. Even if the detector array were very large so that there were plenty of detector elements available to resolve the diffraction pattern, diffraction would still be extremely difficult to observe. The reason is that the diffraction

pattern is most easily seen when the radiation source is monochromatic. Since the emitters are multispectral, not monochromatic, the diffraction pattern must be calculated over the entire wavelength bandpass. Such a calculation would be very difficult, if not impossible, because the relative intensities of the different wavelengths would be governed by a Planck blackbody distribution, which in turn would depend on the temperature of the emitter. Nevertheless, an important generalization can be made. If the bandpass is large enough to include a doubling of the wavelength, then there will be no discernable diffraction pattern. Doubling the wavelength of the radiation will slightly more than double the diameters of the diffraction rings. Thus, the bright rings of one wavelength will overlap with the dark rings of another wavelength.

The multispectral nature of the radiation is the primary reason that diffraction will not be observed, and it explains the recent experimental observation that the shape of the blur "is not Gaussian or Airy–diffraction like, but looks more like a narrow Gaussian added to a broader Gaussian."[7] This description fits the expected appearance of a multispectral diffraction pattern. More work needs to be done in this area. Clearly, the weighting function and the present model for calculating nonuniformity would be improved with a function that better describes the true distribution of the radiation incident on the FPA.

3 Results

The goal of this section is to present the results of our calculations using the statistical model as represented by Eq. (10) and the weighting function, Eq. (12). Unfortunately, this is a daunting task because there are eight different inputs to the model: detector nonuniformity, emitter nonuniformity, sampling ratio, blur size, translation in two directions, and fill factor in two directions. We have attempted to choose those plots that will best represent general trends of relevance to the work going on at the KHILS facility.

First, consider Fig. 5. This plot is made with the same alignment as Fig. 3 and as illustrated for a particular sampling ratio in Fig. 1b. However, this time the full model is used so we have selected a detector nonuniformity of 0.004 and a projector nonuniformity five times larger than that. The major difference, though, is that optical blur is now taken into account. Clearly, as the blur size increases, the effect of positional nonuniformity greatly diminishes. Blurring reduces positional nonuniformity so much that another plot with a different translation would be very similar to this one.
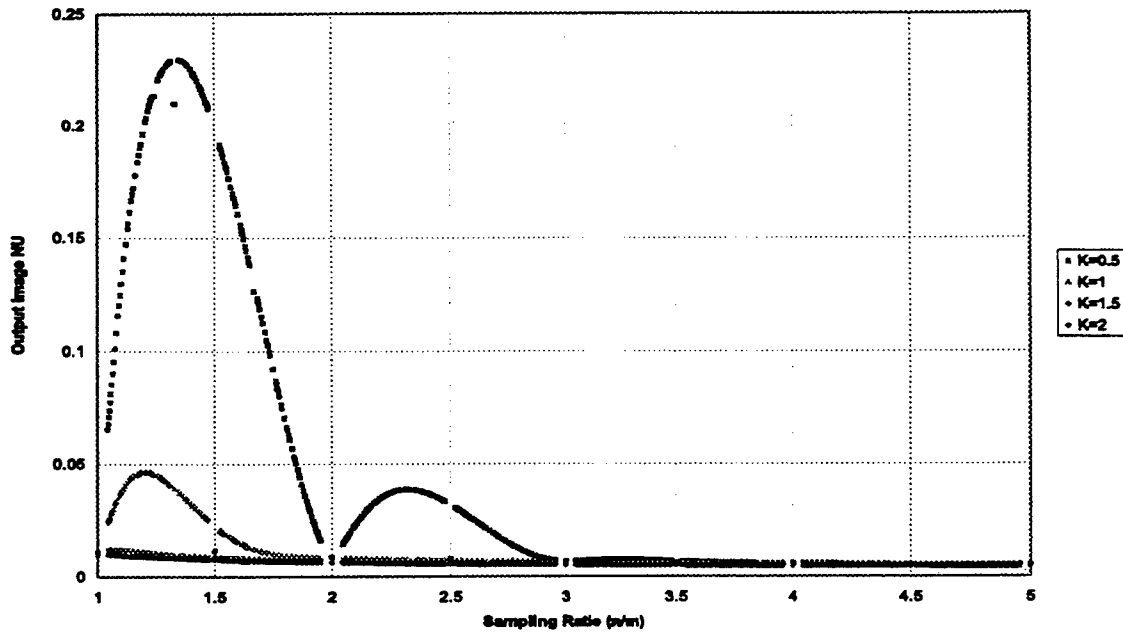
Figure 5: Output NU versus Sampling Ratio with Det. NU = 0.004, Proj. NU = 0.02, $T_x = T_y = 0.0$, and $F_x = F_y = 1$.

Figures 6–12 show output image nonuniformity as a function of projector and detector non-uniformity. The first one, Fig. 6, has an integer sampling ratio of $\frac{n}{m} = 2$. The result is essentially equivalent to our previous work[1] and shows vividly how much more strongly the output depends on detector nonuniformity than emitter nonuniformity because of the weighting factor in Eq. (10). Fig. 7 is unchanged except for choosing a sampling ratio of 2.15. This sampling ratio corresponds to one being used at KHILS during July, 1998, but it is close enough to the integer 2 that the positional nonuniformity does not have much of an effect. Close examination, however, will show that at the origin, where there is no emitter or detector nonuniformity. there is a very small output nonuniformity.

Figure 8 also has a sampling ratio that was typical during July, 1998. Even though $\frac{n}{m} = 1.04$ is quite close to an integer, there is significant positional nonuniformity, which causes the output image nonuniformity to be significant even at the origin. The reason this noninteger sampling ratio is so important is because it falls between the integers 1 and 2, a region of maximum positional nonuniformity (cf. Fig. 5). Figure 9 has the sampling ratio $\frac{4}{3}$ which causes the largest

Figure 6: Output NU versus Detector NU and Projector NU with $n/m = 2$, $K = 1$, $T_x = T_y = 0.0$, and $F_x = F_y = 1$.

positional nonuniformity possible. Figures 10–12 were created under the same circumstances as Figures 7–9 except that the blur size has been increased from $K = 1$ to $K = 1.5$.

Figures 13 and 14 plot output image nonuniformity as a function of translation in the diagonal direction ($T_x = T_y$) for a few different sampling ratios. The first plot has the smaller blur; obviously it has more nonuniformity. Interestingly, the translation can make a big difference for certain sampling ratios. Incidentally, the sampling ratio of 1.5 which causes the large arch in the first figure also causes an arch in the second. However, in the second plot the height of the arch is so little that it does not really show up on this scale.

Figures 15–18 show the output nonuniformity as a function of blur size for different fill factors. The fill factors indicated are the fraction of fill in one direction, but the fill is being changed in both directions simultaneously (eg. $F = 0.9$ is an area fill factor of 81%). Each successive graph is for a greater sampling ratio. The first and last graphs have sampling ratios that were being used in the lab when this work was done. Figure 16 was chosen because its sampling ratio causes maximum positional nonuniformity and Fig. 17 was chosen because it is

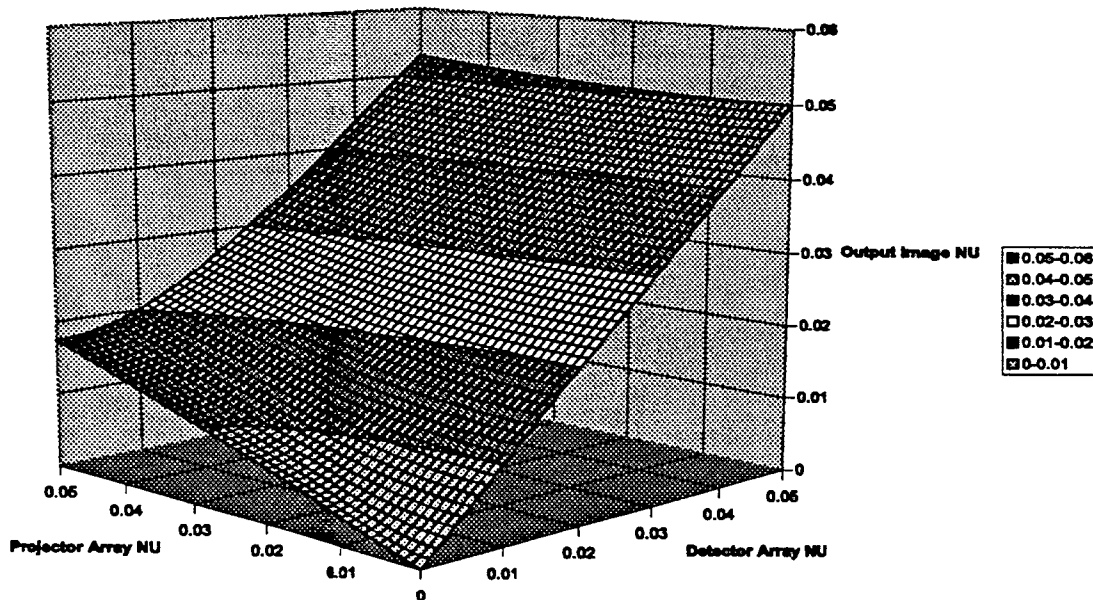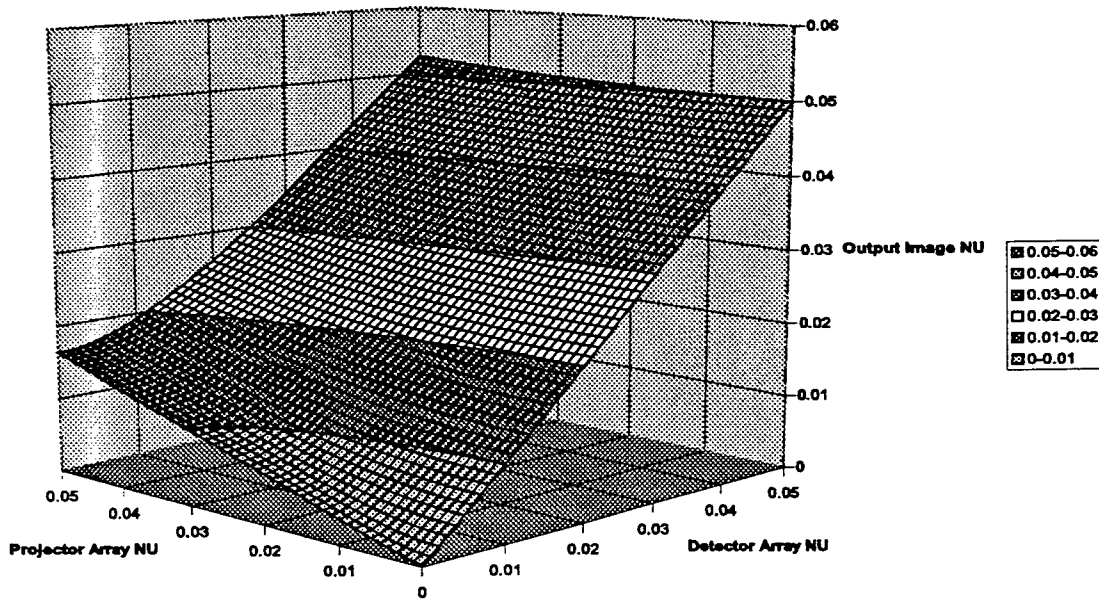Figure 7: Output NU versus Detector NU and Projector NU with $n/m = 2.15$, $K = 1$, $T_x = T_y = 0.0$, and $F_x = F_y = 1$.
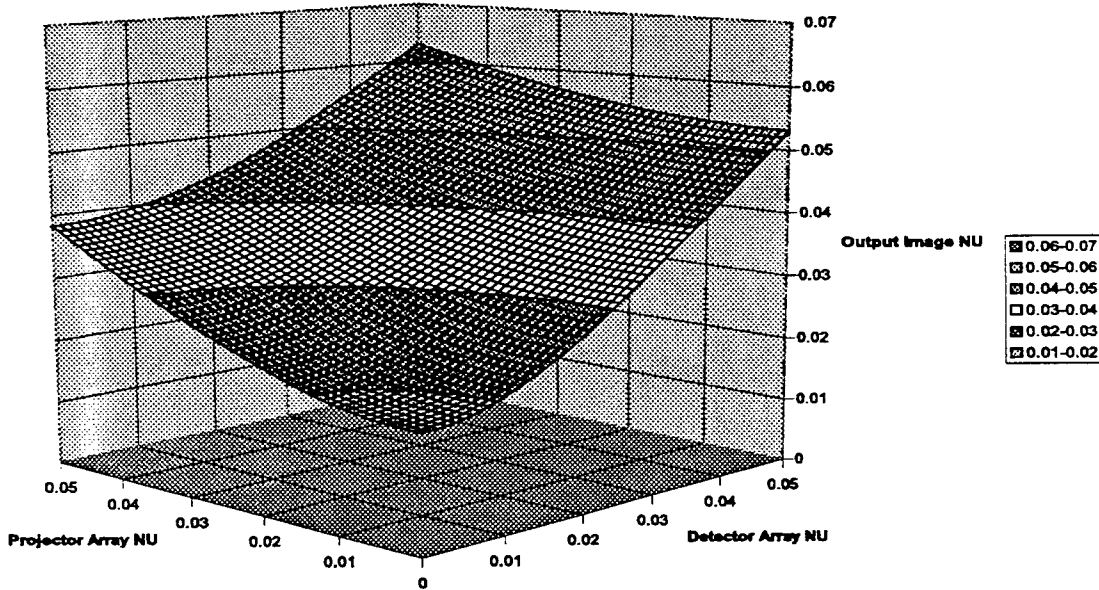
Figure 8: Output NU versus Detector NU and Projector NU with $n/m = 1.04$, $K = 1$, $T_x = T_y = 0.0$, and $F_x = F_y = 1$.
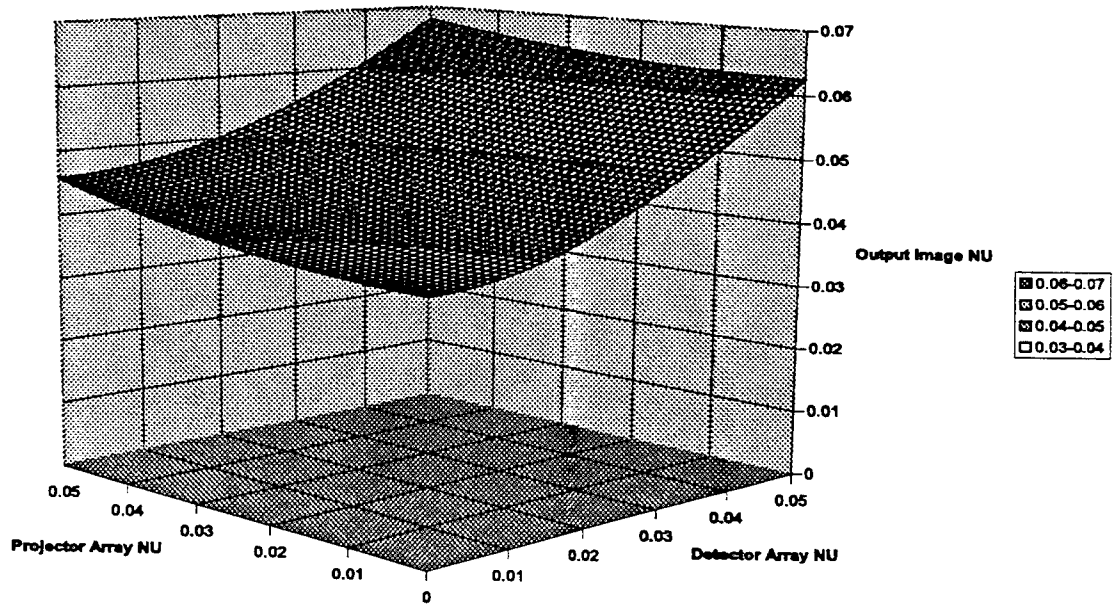
Figure 9: Output NU versus Detector NU and Projector NU with $n/m = 1.333$, $K = 1$, $T_x = T_y = 0.0$, and $F_x = F_y = 1$.
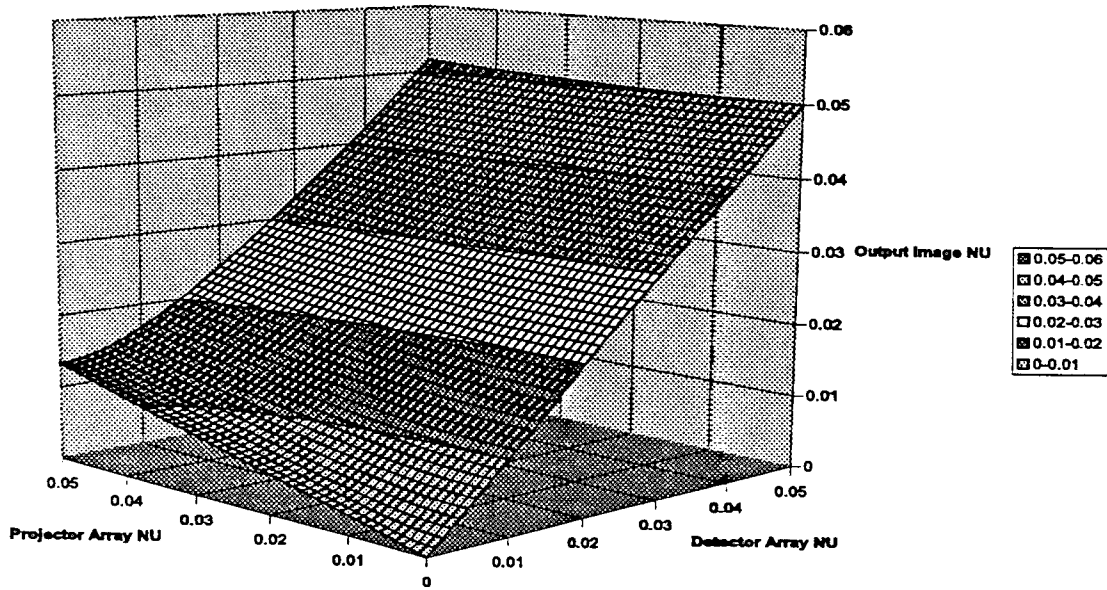
Figure 10: Output NU versus Detector NU and Projector NU with $n/m = 2.15$, $K = 1.5$, $T_x = T_y = 0.0$, and $F_x = F_y = 1$.
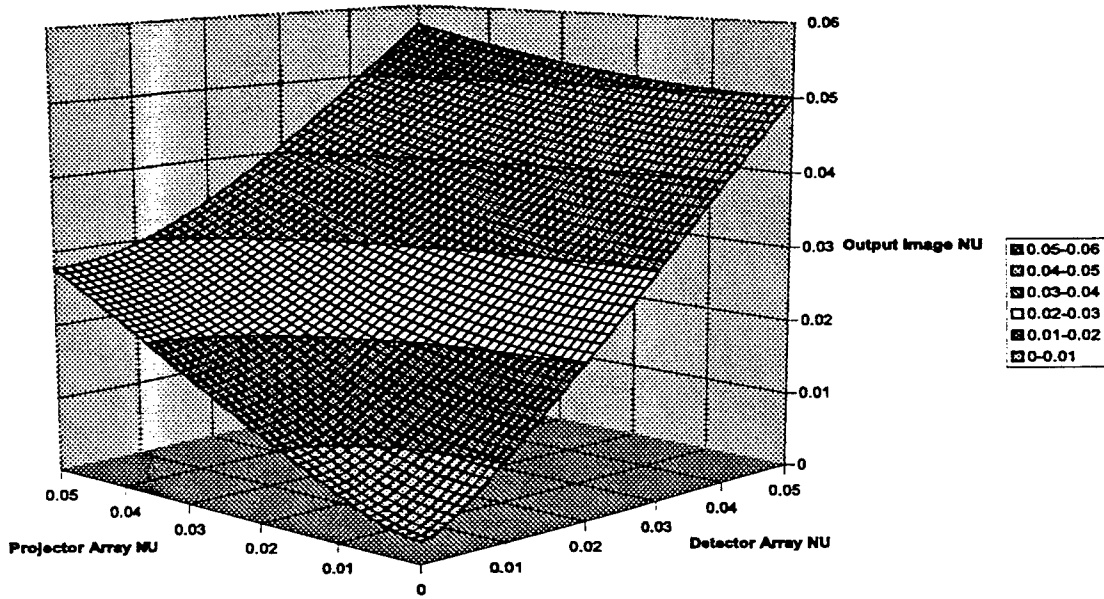
Figure 11: Output NU versus Detector NU and Projector NU with $n/m = 1.04$, $K = 1.5$. $T_x = T_y = 0.0$, and $F_x = F_y = 1$.
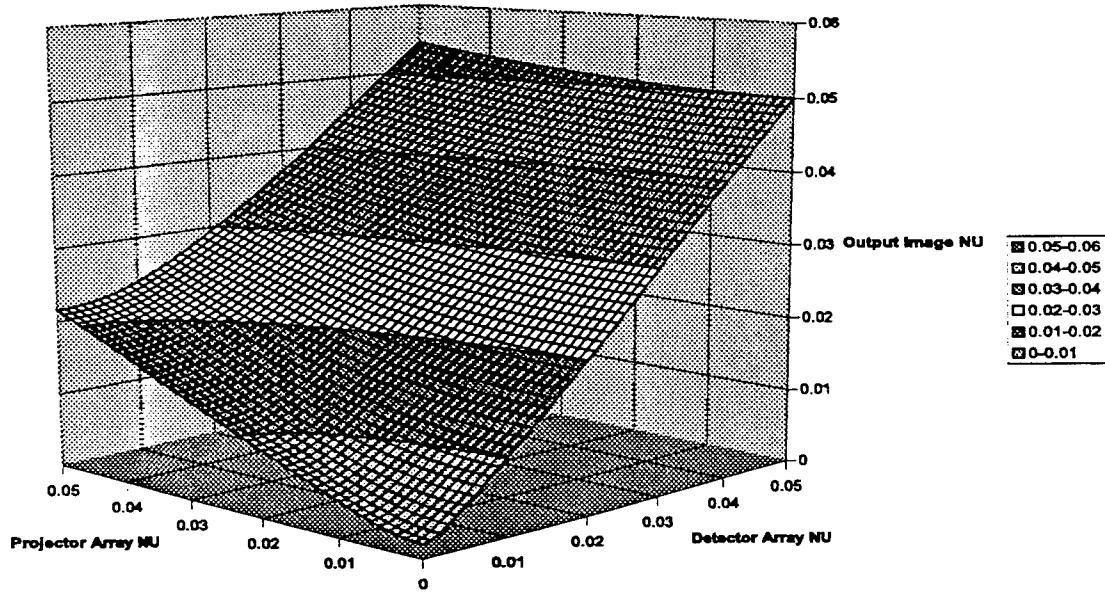
Figure 12: Output NU versus Detector NU and Projector NU with $n/m = 1.333$, $K = 1.5$. $T_x = T_y = 0.0$, and $F_x = F_y = 1$.
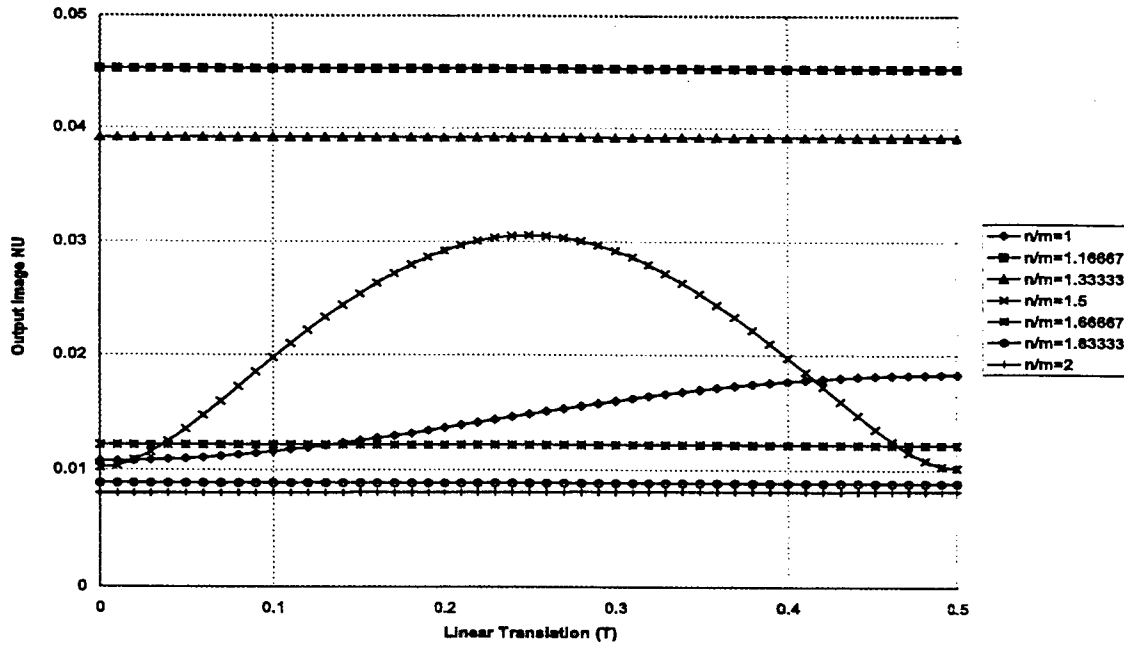
Figure 13: Output NU versus Linear Translation with Det. NU = 0.004, Proj. NU = 0.02, $K = 1$, $F_x = F_y = 1$.
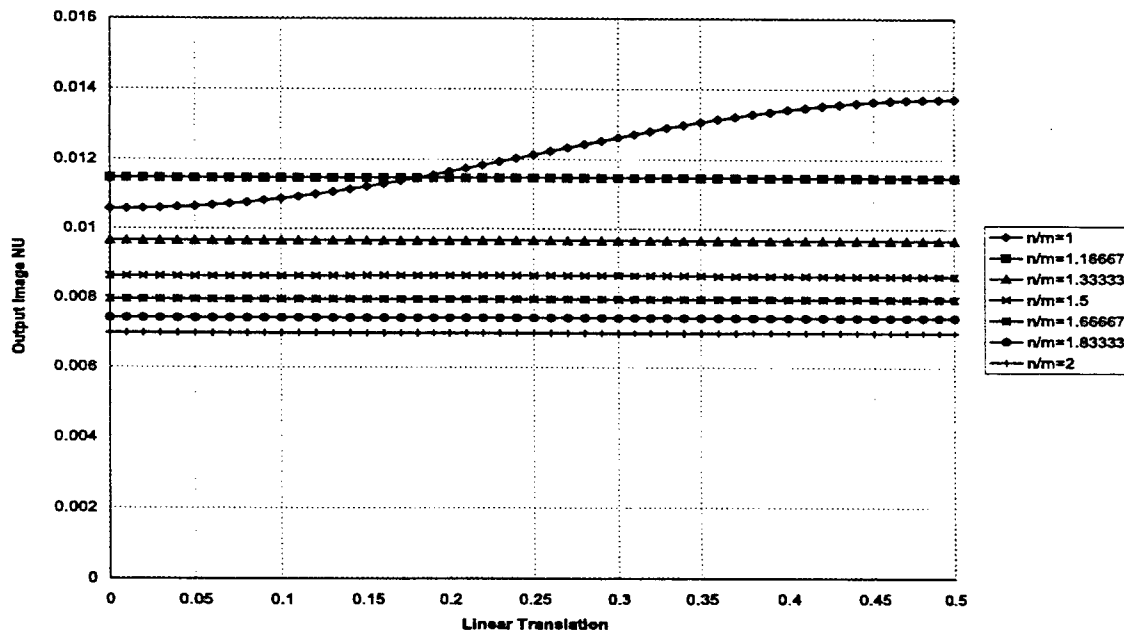
Figure 14: Output NU versus Linear Translation with Det. NU = 0.004. Proj. NU = 0.02, $K = 1.5$, $F_x = F_y = 1$.
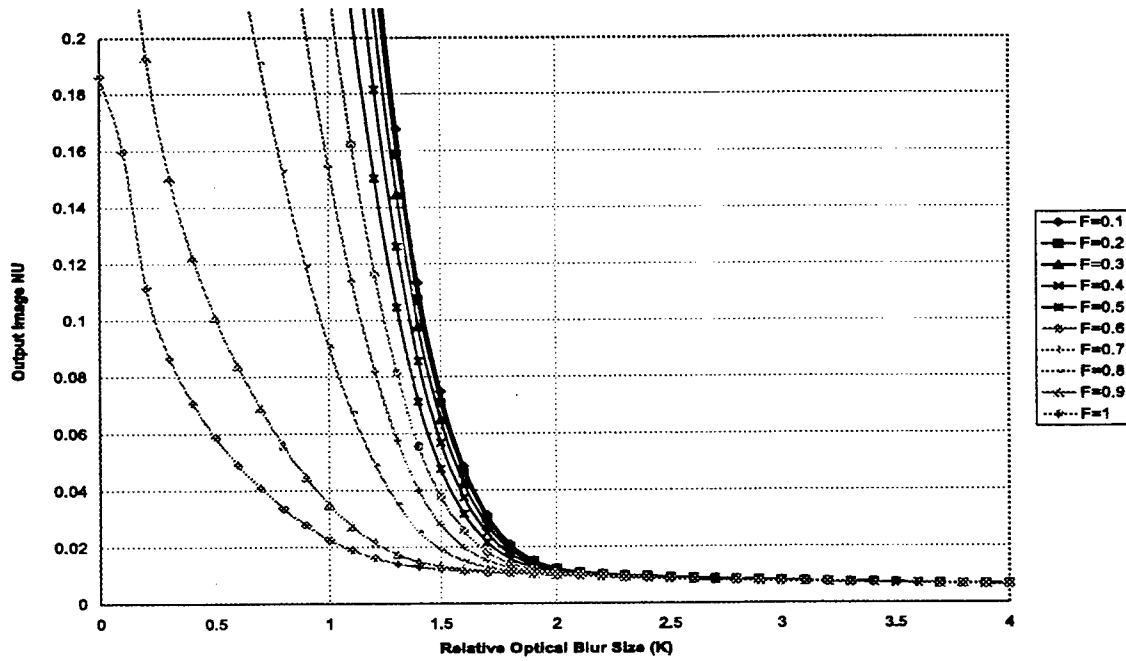
Figure 15: Output NU versus Relative Optical Blur Size with Det. NU = 0.004, Proj. NU = 0.02, $n/m = 1.04$, $T_x = T_y = 0$.
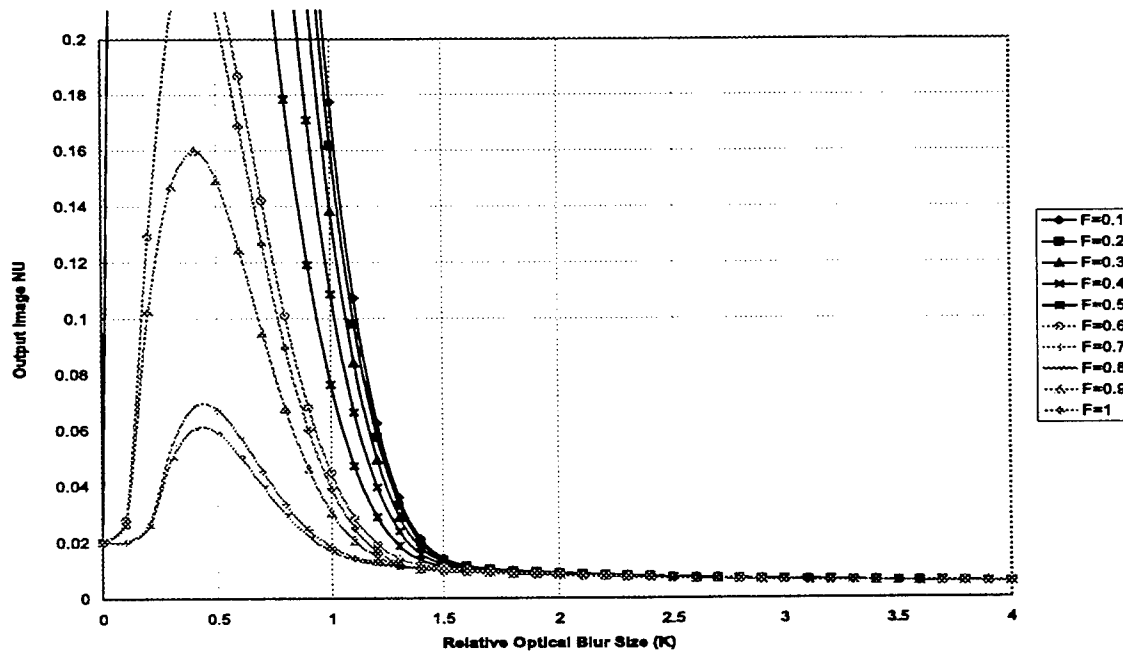
Figure 16: Output NU versus Relative Optical Blur Size with Det. NU = 0.004, Proj. NU = 0.02, $n/m = 1.333$, $T_x = T_y = 0$.
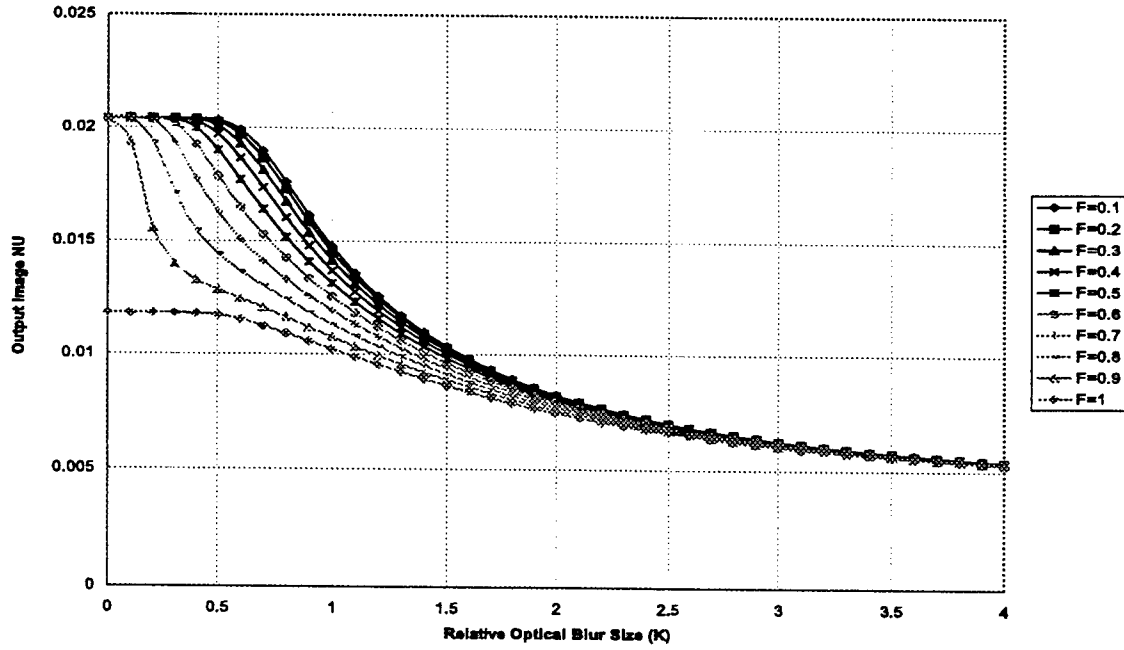
Figure 17: Output NU versus Relative Optical Blur Size with Det. NU = 0.004, Proj. NU = 0.02, $n/m = 1.5$. $T_x = T_y = 0$.
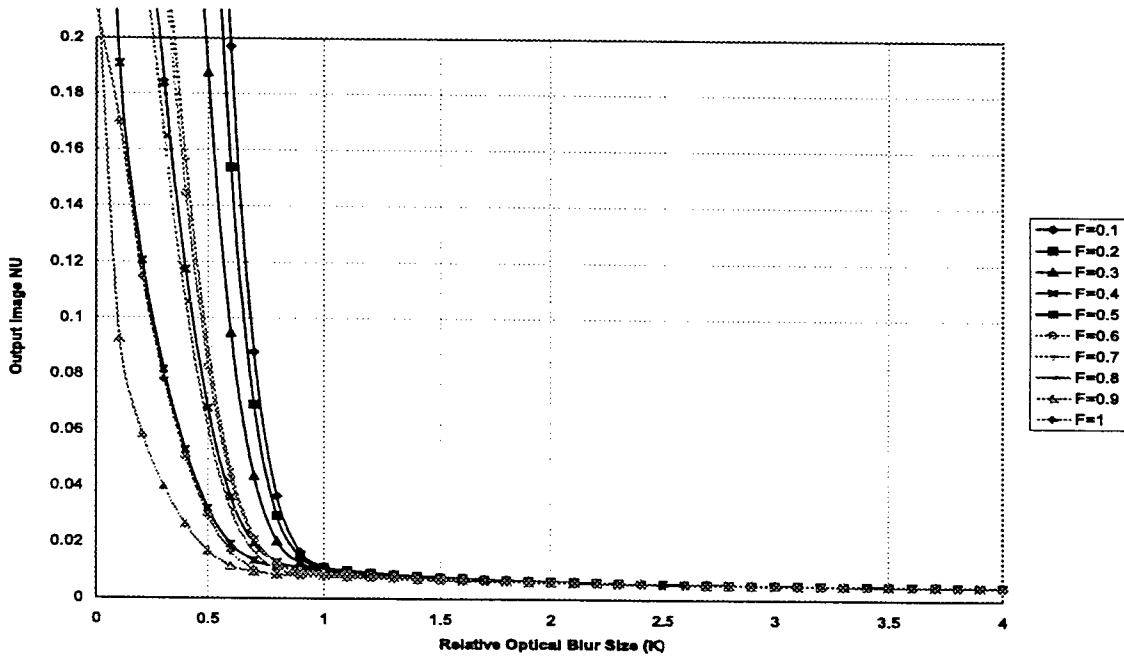
Figure 18: Output NU versus Relative Optical Blur Size with Det. NU = 0.004, Proj. NU = 0.02, $n/m = 2.15$, $T_x = T_y = 0$.
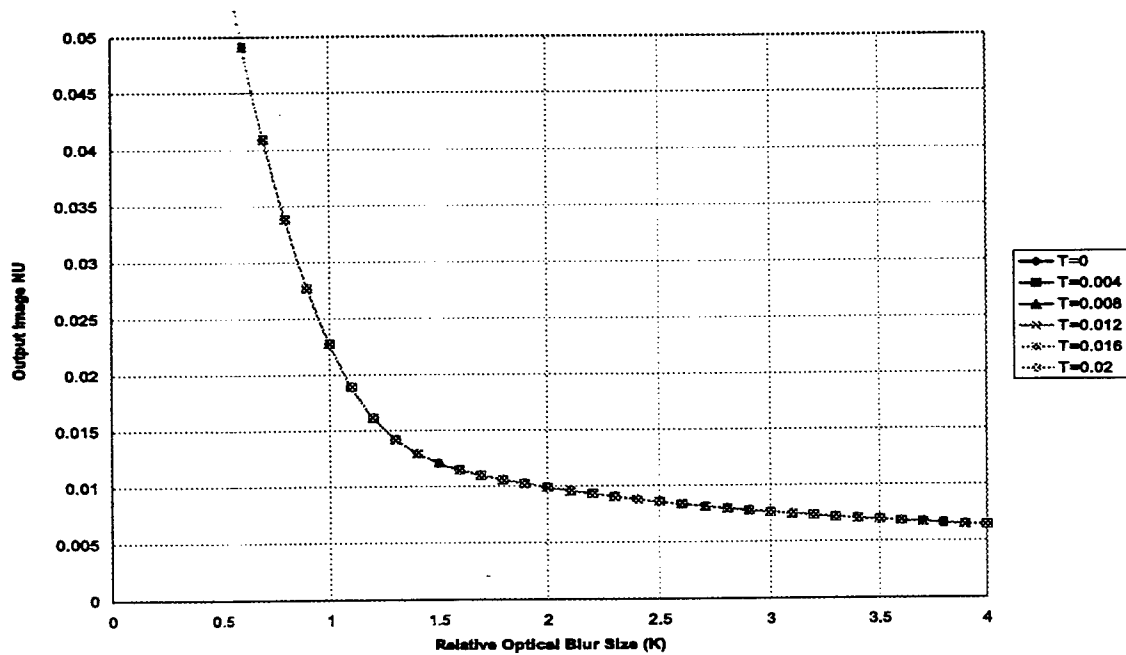
Figure 19: Output NU versus Relative Optical Blur Size with Det. NU = 0.004, Proj. NU = 0.02. $n/m = 1.04$, $F_x = F_y = 1$. All translation lines overlap.
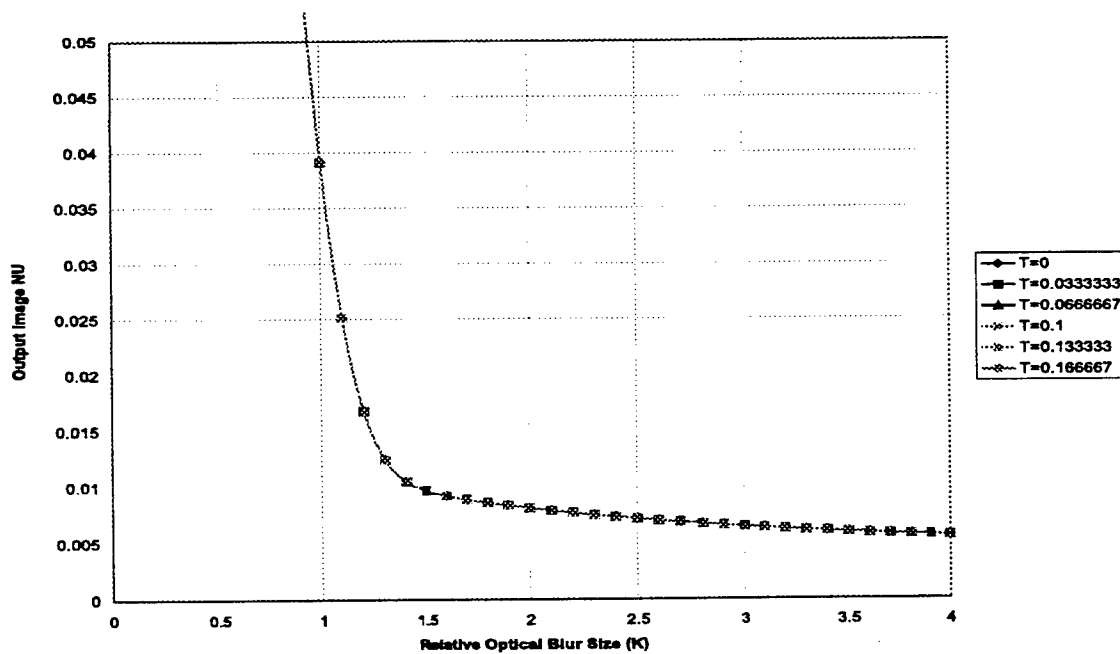
Figure 20: Output NU versus Relative Optical Blur Size with Det. NU = 0.004, Proj. NU = 0.02, $n/m = 1.333$. $F_x = F_y = 1$. All translation lines overlap.

1-23

the one noninteger sampling ratio in this range (for this alignment) which has no positional nonuniformity. It should be reminiscent of last year's work. Figures 16 and 17 are both a little surprising in their behavoir at small blur. The reason is that at the no-blur limit, emitters effectively drop off the edge of the active area of the detectors. Many times this results in one emitter being seen by only one detector—thus the nonuniformity becomes that of the emitter array (which in this case dominates the detector array nonuniformity).

Figures 19 and 20 show output nonuniformity as a function of blur size for different translations. The two sampling ratios chosen for these two graphs happen to be ones that cause translation to have very little effect. Each would appear as one of the straight lines in Figs. 13 or 14. Nevertheless these plots are useful, partly because they provide a different pespective, and partly because they again show the effect of positional nonuniformity. Figure 20 curves upward more steeply at small blur because that particular sampling ratio causes more positional nonuniformity. However, at larger blur sizes, where positional nonuniformity is not much of a factor, it has slightly less nonuniformity because the sampling ratio is higher.

4 Experimental Concerns

Due to the nature of a model or a calculation, there are always going to be questions regarding its relevance to experimental results. This section attempts to address those issues that are most likely to cause concern.

4.1 Accuracy of the Measurement of the Projector's Nonuniformity

Experimentally, the determination of a projector's nonuniformity is a difficult task. Therefore, of the parameters that are entered into the statistical model, the projector's nonuniformity $\frac{\sigma_\phi}{\phi}$ probably has the largest error. How much uncertainty does this create in the calculated test article's output nonuniformity $\frac{\sigma_v}{v}$? An equation predicting this contribution can be found using standard error propagation techniques:

$$\sigma^2_{\left(\frac{\sigma_v}{v}\right)} = \left[\frac{\partial \left(\frac{\sigma_v}{v}\right)}{\partial \left(\frac{\sigma_\phi}{\phi}\right)} \right]^2 \sigma^2_{\left(\frac{\sigma_\phi}{\phi}\right)}. \tag{14}$$

Using Eq. (10) and doing the partial derivative yields

$$\sigma^2_{\left(\frac{\sigma_\psi}{v}\right)} = \frac{\left(\frac{\sigma_\phi}{\phi}\right)^2 \left(\frac{\overline{\sum \omega_{\gamma\delta}^2}}{\overline{\sum \omega_{\gamma\delta}}^2}\right)^2}{\left(\frac{\sigma_{g\eta}}{g\eta}\right)^2 + \left(\frac{\sigma_{\sum \omega_{\gamma\delta}}}{\sum \omega_{\gamma,\delta}}\right)^2 + \left(\frac{\sigma_\phi}{\phi}\right)^2 \frac{\overline{\sum \omega_{\gamma\delta}^2}}{\overline{\sum \omega_{\gamma\delta}}^2}} \; \sigma^2_{\left(\frac{\sigma_\phi}{\phi}\right)}, \qquad (15)$$

which reduces to

$$\sigma_{\left(\frac{\sigma_\psi}{v}\right)} = \frac{\frac{\sigma_\phi}{\phi}}{\frac{\sigma_\psi}{v}} \left(\frac{\overline{\sum \omega_{\gamma\delta}^2}}{\overline{\sum \omega_{\gamma\delta}}^2}\right) \sigma_{\left(\frac{\sigma_\phi}{\phi}\right)}. \qquad (16)$$

It is logical to assume that the average emittance $\overline{\phi}$ is well-known compared to the uncertainty σ_ϕ. If that is true to the extent that all the uncertainty is in the measurement of the uncertainty, then

$$\sigma_{\left(\frac{\sigma_\phi}{\phi}\right)} \approx \frac{\sigma_{\sigma_\phi}}{\overline{\phi}}. \qquad (17)$$

Substituting this approximation into Eq. (16) gives a simpler form,

$$\sigma_{\left(\frac{\sigma_\psi}{v}\right)} = \frac{\frac{\sigma_{\sigma_\phi}}{\sigma_\phi} \left(\frac{\sigma_\phi}{\phi}\right)^2 \frac{\overline{\sum \omega_{\gamma\delta}^2}}{\overline{\sum \omega_{\gamma\delta}}^2}}{\frac{\sigma_\psi}{v}}. \qquad (18)$$

Thus, if the uncertainty in the projector nonuniformity can be estimated, then its effect on the accuracy of this model can be calculated.

4.2 The Point Emitter Approximation

Perhaps the greatest discrepancy between this model and the experimental situation exists because of the use of point emitters in the calculation. In reality the emitters are extended sources. The most recent published description of a WISP array lists a 30% fill factor for the emitters.[8] Obviously, this is not quite a point source, but then neither is it a 100% fill factor array.

It would seem obvious that for this model the difference between a point source and a 30% fill factor extended emitter will make essentially no difference when blur sizes of $K \geq 2$ are used. Since the sampling ratio is always at least unity, $K = 2$ means that the center of the image of a point source will be positioned within the Airy disks resulting from its neighboring emitters. The true extended nature of the emitters will be obscured by the blur. As the blur size is reduced, however, the difference between an extended emitter and a point source will become more important, but larger sampling ratios will still minimize the difference. When $K \leq 1$, the

1-25

Airy disk will be smaller than the detector spacing, so for sampling ratios approaching unity, the difference could be significant.

The main implication of the point source approximation is that it makes the model give values for nonuniformity which must be worst case. Clearly, the more isolated the emitters, the greater the chances of nonuniformity. The difference between extended emitters and point emitters will show up primarily in the positional nonuniformity. If the emitters represented by dots in Fig. (1) were larger, then the radiation would be distributed more evenly between the detector elements. This reduction in the positional nonuniformity is very similar to that observed as blur size is increased.

The use of a point source approximation is not an inherent limit to the approach used by this model. With some effort, it should be possible to incorporate extended emitters in the model. Each emitter would have to be broken into a series of point emitters in much the same manner that the detector elements are already divided into detection points.

4.3 Spatial Droop

Spatial droop (also called busbar robbing) is the term used for the variation in flux from emitters due to a variation in voltage from one emitter to another. The voltage variation is caused by generating bright scenes which require a great deal of current and thus cause a voltage droop across the array.[9] A study of how spatial droop would affect this model has been initiated. The results are interesting and so far seem to be reasonable. However, these thoughts should be considered preliminary at this point.

The goal of this section is to analyze how the nonuniformity due to spatial droop is combined with other sources of nonuniformity. If these calculations are correct, then experimental data can be used to estimate the variation of the voltage. It should also prove possible to subtract from the experimentally measured nonuniformity that part of the nonuniformity caused by a gentle spatial droop across the array.

The power dissipated from a resistor R with a voltage V applied across it is $\frac{V^2}{R}$. If the emitter is approximately a blackbody, this power equals the area of the resistor times σT^4, where T is the temperature in kelvins and σ is the Stefan-Boltzmann constant.

The number of photons emitted within the detector's bandwidth during the integration time

is

$$c = Bf\frac{V^2}{R}, \tag{19}$$

where B is a constant which converts from total dissipated power to photon flux, and f is the fraction of the flux within the bandwidth of the detector. If the emitter is a blackbody,

$$f = \frac{2\pi c^2 h}{\sigma T^4} \int_{\text{bandwidth}} \frac{d\lambda}{\lambda^5 \left(e^{\frac{hc}{\lambda k T}} - 1\right)}. \tag{20}$$

Since we want to distinguish changes due to variations in voltage from fluctuation in the resistances themselves, it is helpful to write

$$R = R_o - \left.\frac{\partial R}{\partial V}\right|_{V_o} (V - V_o), \tag{21}$$

where the resistor has resistance R_o at V_o, and it is assumed that the voltage variations are small enough that only the linear term is needed. Inserting Eq. (21) into Eq. (19) yields

$$\phi = Bf\frac{V^2}{R_o + \left.\frac{\partial R}{\partial V}\right|_{V_o} (V - V_o)}. \tag{22}$$

The standard error propagation equation can now be used:

$$\sigma_\phi^2 = \left(\frac{\partial \phi}{\partial R_o}\right)_{V_o}^2 \sigma_{R_o}^2 + \left(\frac{\partial \phi}{\partial V}\right)_{R_o}^2 \sigma_V^2. \tag{23}$$

First, we need to calculate the partial derivatives:

$$\left.\frac{\partial \phi}{\partial R_o}\right|_{V_o} = \frac{-BfV^2}{R_o^2} = \frac{-\phi}{R_o} \tag{24}$$

and, taking advantage of Eq. (22) along with a little algebra,

$$\left.\frac{\partial \phi}{\partial V}\right|_{R_o} = \left.\frac{\partial f}{\partial V}\right|_{V_o} \frac{\phi}{f} + 2\frac{\phi}{V} - \left.\frac{\partial R}{\partial V}\right|_{V_o} \frac{\phi}{R_o}. \tag{25}$$

Putting these results into Eq. (23), yields

$$\left(\frac{\sigma_\phi}{\phi}\right)^2 = \left(\frac{\sigma_{R_o}}{R_o}\right)^2 + \sigma_V^2 \left[\frac{\left.\frac{\partial f}{\partial V}\right|_{V_o}}{f} + \frac{2}{V_o} - \frac{\left.\frac{\partial R}{\partial V}\right|_{V_o}}{R_o}\right]^2. \tag{26}$$

In this equation, the first term on the right, $\left(\frac{\sigma_{R_o}}{R_o}\right)^2$, represents the portion of $\left(\frac{\sigma_\phi}{\phi}\right)^2$ due to the variation in emitter resistances. The nonuniformity of the projector array is normally understood to be this term. The second term in Eq. (26) accounts for the nonuniformity caused by spatial droop.

1-27

Marlow et al.[10] used an Amber AE4128 camera to analyze an emitter array. The camera had a 3–5μm bandwidth, and the apparent temperatures used were from approximately 300K to 350K. They state that when all 256 × 256 emitters were turned on to maximum value, the peak-to-valley difference divided by the average radiance was less than 10%.

If the emitter is acting as a blackbody, then σT^4 is proportional to $\frac{V^2}{R}$, so a 1% change in V produces a 0.5% change in T. At 350K, a 1% change in T (2% change in V) produces approximately a 5.5% change in f. Therefore, near 350K, for a blackbody emitter,

$$V_o \frac{\frac{\partial f}{\partial V}\big|_{V_o}}{f} \approx \frac{\frac{\Delta f}{f}}{\frac{\Delta V}{V_o}} = \frac{5.5\%}{2\%} = 2.75 \tag{27}$$

Therefore, near 350K,

$$\left(\frac{\sigma_\phi}{\phi}\right)^2 = \left(\frac{\sigma_{R_o}}{R_o}\right)^2 + \left(\frac{\sigma_V}{V_o}\right)^2 \left[2.75 + 2 - \frac{V_o}{R_o}\left(\frac{\partial R}{\partial V}\right)_{V_o}\right]^2 \tag{28}$$

The third term in the square brackets $\frac{V_o}{R_o}\left(\frac{\partial R}{\partial V}\right)_{V_o}$ depends on the coefficient of linear expansion and the temperature coefficient of resistivity of the material. For the silicon-nitride wafer resistors described in Ref. 8, this term should be negligible compared with the other terms in the brackets, 2.75 and 2. Therefore,

$$\left(\frac{\sigma_\phi}{\phi}\right)^2 \approx \left(\frac{\sigma_{R_o}}{R_o}\right)^2 + \left(\frac{\sigma_V}{V_o}\right)^2 [4.75]^2. \tag{29}$$

This equation implies that the 10% variation in radiance reported in Ref. 10 would have resulted from approximately a 2% variation in the voltages across the array.

This result becomes important to the statistical model when scenes that are predominately bright are being projected by the emitter array. In such a case, a gradual droop is expected across the projector array. This droop will show up as a large nonuniformity in the standard measurements. However, a gradual droop will have little impact on the projector's ability to display fine features with the appropriate contrast for testing.

Equation (29) provides a method of correcting the measured nonuniformity $\frac{\sigma_\phi}{\phi}$ for spatial droop effects because the final term represents the nonuniformity due to droop. Only the $\left(\frac{\sigma_{R_o}}{R_o}\right)^2$ term really matters for the determination of nonuniformity which will degrade the simulation. This resistor nonuniformity can be calculated by subtracting the last term from the measured nonuniformity. The problem with this technique is that $\frac{\sigma_V}{V_o}$ is not likely to be known. An equation could, however, be written that would model the voltage droop across the array and then the nonuniformity of the voltage could be calculated.

4.4 Rotation

Some preliminary effort has been put into considering the effects of rotating the FPA relative to the projector array. So far it seems that rotation will effect both positional nonuniformity and the contribution of emitter array nonuniformity. It does look like the model can be successfully made to include rotational effects. It is clear, however, that computational time will skyrocket.

Conclusion and Recomendations

The 1997 statistical model[1] has been sucessfully expanded and refined to account for non-integer sampling ratios. During this work, the phenomenon of positional nonuniformity was identified, and a method for quantifying it was devised. Positional nonuniformity exists independent of detector and emitter nonuniformity. It is caused by different detector elements "seeing" the images of different numbers of emitters. The variation from detector to detector occurs in a systematic and predictable manner with a two-dimensional spatial period we have chosen to call a chunk. Positional nonuniformity occurs for virtually all non-integer sampling ratios. but decreases dramatically as integer sampling ratios are approached. Furthermore, the general trend in positional nonuniformity is smaller as the sampling ratio is increased, despite large local maxima between integer values. As the optical blur is increased, the effects of positional nonuniformity decrease significantly. However, the model currently does predict significant nonuniformity in at least part of the optical blur range applicable to the actual experimental situation.

All of the other issues in the proposal[5] for this effort have been addressed. Section 1.3 discusses the concerns about offset nonuniformity and hopefully puts them to rest. A method for determining how the uncertainty of the model's output depends on the uncertainty of the important and hard to determine emitter nonuniformity is derived in Section 4.1. Significant progress has been made in identifying a means of handling spatial droop in the context of this model. In fact, the results of Section 4.3 provide a method for separating the effects of different emitter resistances from the variation in voltage applied to them because of droop. The last issue the proposal addressed was the effects of rotation of the detector array relative to the emitter array. This issue is still open, but enough preliminary work was done to determine that the nonuniformity will be affected by rotation and that quantifying its effect should be possible

as a logical extension of the current model.

As part of this research, considerable effort was put into a consideration of the effects of diffraction and the relative validity of a two-dimensional Gaussian as an adequate approximation for the optical blur or point spread function. It is readily admitted that the Gaussian is only a rough but convenient approximation. It was also determined that a Fraunhofer diffraction pattern will not be observed, primarily because of the multispectral nature of the radiation and partly because of the limited resolution. For now, the most expedient approach was to continue with the Gaussian approximation, but if this model is to be developed further, then this issue should be revisited.

At this point there seems to be a number of things yet to do to strengthen and extend the current model:

- As discussed above, the use of the Gaussian approximation should be improved upon. Two possible approaches come to mind. First, careful examination of appropriate data could lead to the formulation of a useful phenomenological approximation. This approach has been used at least once, by Flynn et al.[7], and it could be incorporated into the present work. Second, it should be possible to calculate at least approximately what the shape of the blur would be by considering multispectral diffraction in the appropriate bandwidth. Combining these methods would have the advantage of providing a comparison between theory and experimental results, thus offering the opportunity for a more complete understanding of what is going on in this crucial area.

- Probably the most critical area of concern regarding the present model is the use of a point emitter approximation when the real emitters are truly extended. The extended sources will certainly reduce the effects of positional nonuniformity, especially at small blur sizes. It is also true that the fact of extended sources will significantly alter the shape of the blur. Thus, this issue is intimately connected to the previous one. A method for dealing with it in the context of this model is readily apparent. It should be possible to use a series of point emitter approximations to represent one extended emitter. Such an approach would not only improve this model, but should shed significant light on the blur shape issue.

- We got far enough on the rotational work to determine that it should be doable with our approach. We have a high degree of confidence that with more time the model can be ex-

tended to include this effect. We believe it is an important issue because the nonuniformity will be affected and KHILS is planning to rotate the detector arrays during testing. Thus, it is important to know the degree to which the nonuniformity will be affected and which parameters are important.

- Although a lot was accomplished with regard to spatial droop, it would be nice to model how spatial droop occurs and then subtract it from the predicted nonuniformity. Being able to do this would allow low-frequency spatial droop to be removed from the measurement of emitter array nonuniformity. This task is not expected to be difficult.

- We have been told that a good model for nonuniformity would be useful to KHILS as it deals with potential customers. If this is true, it might be useful to have a professional-looking program and interface that would calculate and display the results of the model for particular input parameters. Currently eight different inputs are required, so it is impossible to provide a plot which covers every possible situation. The program is not too difficult to use as it stands; however, the output is a matrix of raw numbers which must be put into another application to produce useful graphs. This situation is likely to limit the use of the program.

Much has been accomplished in the development of a useful model which will predict the non-uniformity of a test article's output image. Although there is much more that can be done, hopefully the present work will prove to be a useful tool.

5   Acknowledgments

References

1. B. P. Beecken and T. J. Belich, "Calculation of the contribution of the non-uniformity of the scene projector to the non-uniformity of the output image of a test article," *Proceedings of SPIE*, vol. 3368, p. 126.

2. D. L. Garbo, E. M. Olson, C. F. Coker, and D. R. Crow, "Real-time Three Dimensional Infrared Scene Generation Utilizing Commercially Available Hardware," The 10th Annual International Aerospace Symposium, April 1996.

3. L. E. Jones, R. G. Stockbridge, A. R. Andrew, W. L. Herald, and A. W. Guertin, "Characterization Measurements of the Wideband Infrared Scene Projector Resistor Array (Part II)," presented at SPIE, Orlando, Florida, April 1997.

4. L. E. Jones, E. M. Olson. R. L. Murrer, and A. R. Andrew, "A Simplified Method for the Implementation of Nonuniformity Correction on a Resistor Array Infrared Scene Projector," presented at SPIE, Orlando, Florida, May 1997.

5. B. P. Beecken, "Development of a Statistical Model Predicting the Impact of a Scene Projector's Nonuniformity on a Test Article's Performance," unpublished research extension grant proposal to AFOSR, October, 1997.

6. W. L. Wolfe and G. J. Zissis, Eds., *The Infrared Handbook*, prepared by the Infrared Information Analysis Center, Ann Arbor, MI, (1989) p. 8-28.

7. D. S. Flynn, S. A. Marlow, and E. M. Olson, "Projector nonuniformity and spatial effects modeling," *Proceedings of SPIE*, vol. 3368, p. 189.

8. B. Cole, R. Higashi, J. Ridley, J. Holmen, R. Stockbridge, L. Murrer, E. Burroughs. "Large-area infrared microemitter arrays for dynamic scene projection," *Proceedings of SPIE*, vol. 3368, p. 57.

9. A. P. Pritchard, M. A. Venables, and D. W. Gough, "Output accuracy and resolution limitations in resistor array infra-red scene projection systems," *Proceedings of SPIE*, vol. 2742, p. 15.

10. S. A. Marlow, D. S. Flynn, and J. R. Kircher, "Radiometric, noise, and spatial characterization of the Wideband Infrared Scene Projector." *Proceedings of SPIE*, vol. 3368, p. 157.

IMPLEMENTATION OF AN OPTIMIZATION ALGORITHM IN ELECTROMAGNETICS FOR RADAR ABSORBING MATERIAL LAYERS

John H. Beggs
Assistant Professor
Department of Electrical and Computer Engineering
Box 9571
Mississippi State, MS 39762

1 Objective

The objective in this effort was to implement a optimization algorithm for multiple Dallenbach layers based upon a genetic algorithm to provide broadband RCS reduction from a planar metal surface. Each Dallenbach layer has a certain thickness, d, and material properties of permittivity, ϵ, and permeability, μ. The layers are placed on the planar metal surface and the parameters are chosen so as to minimize some quantitative measure of the RCS versus frequency and incidence angle. A specific genetic algorithm has been implemented to minimize a RCS cost function with respect to the input parameters. This report provides a description of Dallenbach layers and the appropriate electromagnetic theory along with a brief description of the genetic algorithm.

2 Technical Approach

2.1 Introduction

The technical approach used to solve this problem was based upon a binary genetic algorithm that encoded the Dallenbach layer thickness and material properties in an appropriate chromosome. These individual layer chromosomes were then combined into one large chromosome, which was subsequently used to evaluate the cost function for the next iteration. Convergence was assumed when the cost function varied within a prescribe tolerance between subsequent iterations. The technical approach used in this project was based heavily on a method previously published in the literature [1], and intended to enhance this method by using a different cost function. This complete approach would then be suitable to include in a larger multidisciplinary design optimization (MDO) program for optimizing a flight vehicle for structures, aerodynamics and Radar Cross Section (RCS). This algorithm would provide the RCS optimization of the flight vehicle, given that the small surface patches could be assumed to be locally planar.

The progress to date since contract award includes an operational C++ genetic algorithm code which can be used for optimization of radar absorbing material layers. The code is based upon reflection from a stratified media placed over a perfectly conducting backing (i.e. metal). The genetic algorithm used is a binary genetic algorithm which codes the material type and thickness for each layer in binary numbers. Various sections of the code have been debugged and tested and are working properly. The code has been benchmarked against analytical solutions for several two and three-layer Dallenbach layer configurations at different frequencies. The code was also benchmarked against previously published results in the open literature [1], but several discrepancies were noticed between the previously published results, analytical solutions and the algorithm developed in this project. The previously published results differed from the analytical solutions and since our algorithm agreed with analytical solutions, it is assumed that the previously published results were either mislabeled or somehow incorrect. Future work to continue this project and genetic algorithm development include benchmarking the code against more complicated analytical solution configurations and determining the exact nature of

discrepancies with the previously published results. Extensive convergence studies also need to be performed to determine if a fitness scaling function is required to enhance the convergence of the genetic algorithm. Further enhancements and improvements to the genetic algorithm would also be in order.

2.2 Theory

A Dallenbach layer is a layer of dielectric material that is placed directly on the conducting surface as shown in Figures 2 and 3. The performance of the layer is analyzed
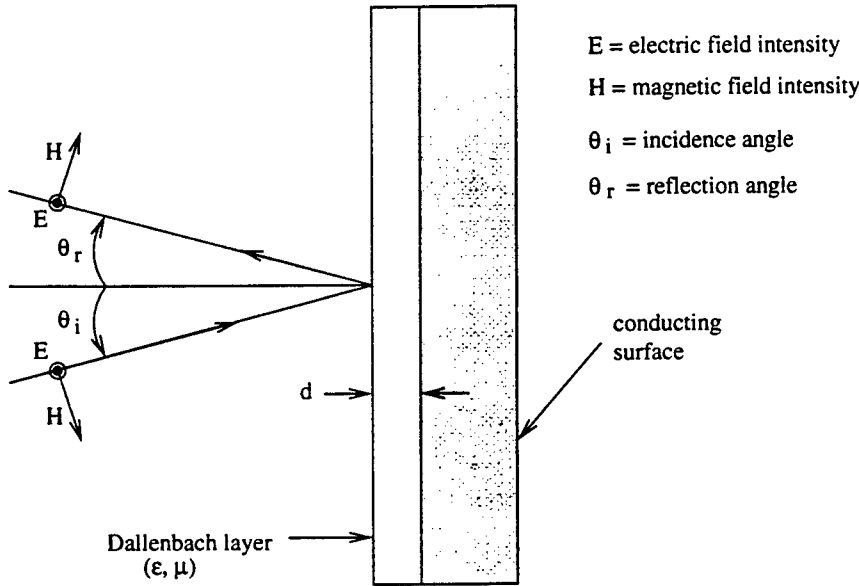
Figure 1: Geometry for single Dallenbach layer for RCS reduction using perpendicular polarization.

in terms of the reflection coefficient for both horizontal and vertical polarizations as a function of the permittivity and permeability of the material as defined by $\epsilon_r = \epsilon' - j\epsilon''$ and $\mu_r = \mu' - j\mu''$, respectively. The thickness of the single Dallenbach layer also affects the reflection coefficient and it can be obtained from the zero reflection condition at the surface of the layer by the expression

$$d = \frac{1}{\gamma} \tanh \left(\frac{Z_0}{Z} \right) \tag{1}$$

where γ is the complex propagation constant in the material layer, Z_0 is the wave impedance of free space, Z is the characteristic wave impedance of the material and is given by

$$Z = X + jY \tag{2}$$

where

$$X = Z_0 \sqrt{\frac{\mu_r' \cos \delta_e}{\epsilon_r' \cos \delta_m}} \cos \frac{\delta_e - \delta_m}{2} \tag{3}$$
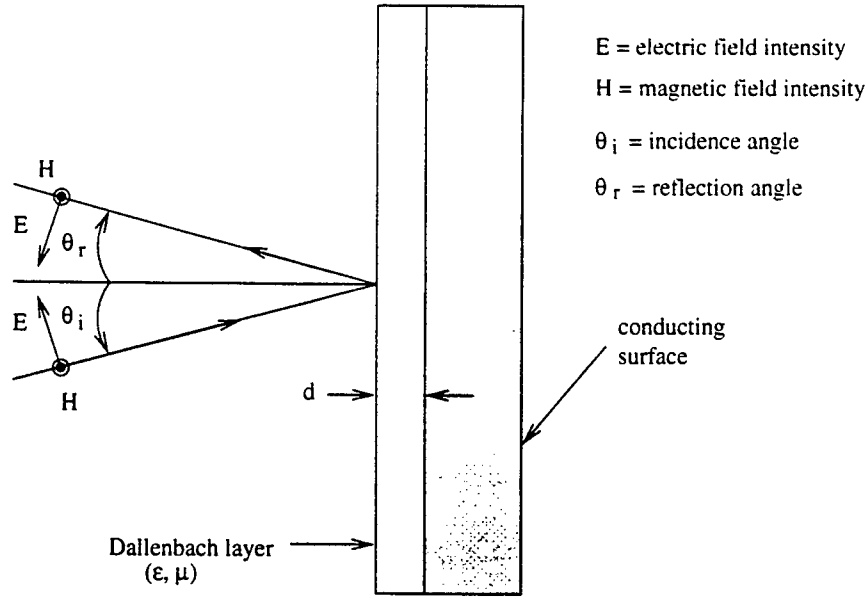
Figure 2: Geometry for single Dallenbach layer for RCS reduction using parallel polarization. Note the difference in the electric and magnetic field vectors from Figure 2.

$$Y = Z_0 \sqrt{\frac{\mu_r' \cos \delta_e}{\epsilon_r' \sin \delta_m}} \cos \frac{\delta_e - \delta_m}{2} \tag{4}$$

and $j = \sqrt{-1}$ and δ_e, δ_m are the electric and magnetic loss tangents, respectively. For a given Dallenbach layer thickness, equations (1), (3) and (4) can be used with an optimization algorithm to obtain the material parameters ϵ_r', μ_r', ϵ_r'' and μ_r'' (since δ_e and δ_m are directly related to ϵ_r'' and μ_r''). Note that there are four degrees of freedom, but constraints can be placed on the variables to make the optimization a tractable problem. The single Dallenbach layer is a frequency sensitive method and works best for normal incidence. It can be designed for off normal incidence if desired. Using multiple Dallenbach layers increases the absorption bandwidth, leading to broadband RCS reduction.

For multiple Dallenbach layers and parallel polarization [2], assume the geometry consists of M layers, each with a thickness, d_m, and with material parameters ϵ_m, μ_m and σ_m. Further assume that the Mth (or last) layer is a semi-infinite layer with $d_M \rightarrow \infty$ so that no reflected waves exist in the last layer (M). The reflection coefficient, $R_\parallel$, is given by the expression

$$R_\parallel = \frac{K_0 - Z_1}{K_0 + Z_1} \tag{5}$$

where

$$Z_m = K_m \frac{Z_{m+1} + K_m \tanh u_m d_m}{K_m + Z_{m+1} \tanh u_m d_m} \tag{6}$$

is the impedance of the m-th layer with

$$K_m = \frac{u_m}{\sigma_m + j\omega\epsilon_m} \tag{7}$$

and

$$u_m = \sqrt{k_0^2 + \gamma_m^2} \tag{8}$$

The term k_0 is the free space wave vector defined by $k_0 = \omega/c$ and the term γ_m is the complex propagation constant for material m defined by

$$\gamma_m = \sqrt{j\omega\sigma_m\mu_m - \epsilon_m\mu_m\omega^2} \tag{9}$$

In the case when the last layer is a conducting surface, the impedance of the final layer is given by

$$Z_M = 0 \tag{10}$$

For the perpendicular polarization, the reflection coefficient is given by

$$R_\perp = \frac{N_0 - Y_1}{N_0 + Y_1} \tag{11}$$

where

$$Y_m = N_m \frac{Y_{m+1} + N_m \tanh u_m d_m}{N_m + Y_{m+1} \tanh u_m d_m} \tag{12}$$

is the admittance of the m-th layer with

$$N_m = \frac{u_m}{j\omega\mu_m} \tag{13}$$

and u_m as defined in (8). For the case in which the last layer is a conducting surface, the admittance of the final layer is given by

$$Y_m \to \infty \tag{14}$$

2.3   Description of Optimization Method

The original plan in the statement of work for this project called for implementation of a projected gradient optimization algorithm for multiple Dallenbach layers to provide broadband RCS reduction from a planar metal surface. However, in a search of the literature, a different approach was identified as a preferred optimization method. This optimization method is a genetic algorithm, which has already been used for various applications in electromagnetics. The genetic algorithm was chosen over the project gradient method for several reasons. A genetic algorithm optimizes with continuous or discrete parameters, it does not require gradient information, it simultaneously searches from a wide sampling of the cost surface, it deals with a large number of parameters and is well suited for parallel computing, it optimizes parameters with complicated cost surfaces, it does not get stuck in a local minimum, it provides a list of optimum parameters, it can work with encoded parameters and, finally, it can work with numerical data, experimental data or analytical functions [3]. A genetic algorithm has already been used to design lightweight, broad-band microwave absorbers suitable for reducing radar signatures [1] and this approach forms the basis for the work. Reference [1] provides a good description of the entire genetic algorithm, and some highlights of that approach will be briefly cited

here.

This genetic algorithm will optimize multiple Dallenbach layers for minimum reflection given a predefined set of frequency-dependent materials. The optimization will simultaneously determine the optimal material choice for each layer and its thickness over a prescribed range of incidence angles and frequencies. An upper bound is placed on the number of layers and the thickness of each layer. The material for each layer is binary coded in a sequence M_j of m bits and the thickness is binary coded in a sequence T_j using t bits. The material choice for each layer, $L_j = M_j T_j$, includes the material and thickness bits. A chromosome, G, in the genetic algorithm is formed by combining the material choices for all layers as $G = L_1 L_2 \ldots L_{N_l}$. The flowchart for the genetic algorithm approach is shown in Figure 3. This flowchart represents a typical genetic algorithm with
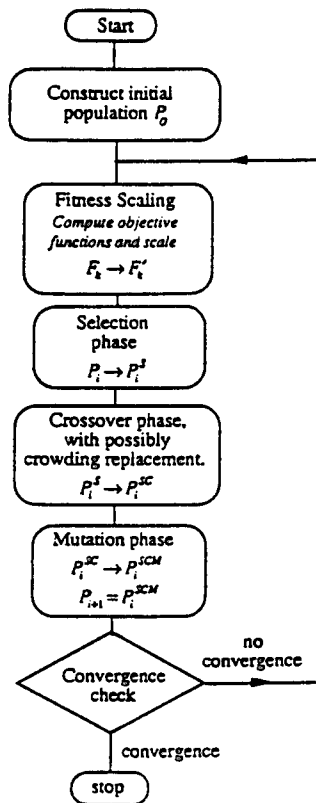
Figure 3: Flowchart for proposed genetic algorithm to optimize a series of Dallenbach layers over a planar metal surface for reduced radar signature.

a selection phase, a crossover (or "mating") phase and a mutation phase. The details of each phase given in [1] are quite lengthy, and will be omitted here for the sake of brevity. However, a good description of the various components of this particular genetic algorithm are given in [1] and descriptions for more generic genetic algorithms are given in [3].

The initial population, P_0, is filled by constructing random sequences of bits for each chromosome in P_0. The initial population is then sorted and any clones are replaced by random sequences until all initial sequences are unique. The top one-half of the initial population, P_0, then becomes the iterative (or working) population, P_i. This population is subsequently used during the iterative portion of the optimization algorithm.

The crossover (or mating) phase uses the top 1/2 of the iterative population of the genetic algorithm. Chromosomes are randomly paired and a random crossover bit (other than 0) is selected for mating. The child sequences are obtained from the parent sequences by combining the group of bits to the left of the crossover bit of the first parent with the group of bits to the right of the crossover bit of the second parent, and vice versa [?, michielssen93a] Any zero child sequences are replaced with a random sequence, and any clones are replaced with random sequences to add further versatility to the optimization procedure.

At the present time, a mutation phase has not been implemented in the optimization procedure. It is, however, a simple matter to add this feature to the current optimization algorithm. As mentioned in [1], a fitness scaling function was employed so that the genetic algorithm would not lead to premature convergence due to a few highly fit sequences in the initial population, or from poor convergence due to a large number of fit, but very similar, sequences. This fitness scaling function was also not implemented in the present algorithm, but could be implemented in the future after extensive convergence studies have been performed. Convergence is achieved when the difference in the mean of the cost function from one iteration to the next is less than a prescribed tolerance level.

The cost function used in this optimization procedure is based upon the reflection coefficient. The genetic algorithm optimizes a cost function related to the reflection coefficient over several incidence angles (θ) and frequencies (f) of interest. The actual cost function is given by

$$f = max_{i,j} \left[1 - \left| R^{\|/\perp} \left(\theta_i, f_j \right) \right| \right] \tag{15}$$

where the i and j denote individual incidence angles and frequencies, and R is the reflection coefficient for both parallel ($\|$) and perpendicular ($\perp$) polarizations. The cost function is averaged over all incidence angles and frequencies to determine the mean cost, which is the final value of the cost function used for optimization.

2.4 Results

To benchmark the genetic algorithm code against analytical solutions, several simple Dallenbach layer problems were constructed for the parallel polarization to verify the cost function calculation portion of the optimization procedure. The first case is a single Dallenbach layer over a PEC backing. The material layer is 2 mm thick, with material properties $\epsilon_r = 10 + j0$, $\mu_r = 1$, and the frequency is 1 GHz. The analytical reflection coefficient is $\gamma_{exact} = 0.996 - j0.084$ and the genetic algorithm code produces the result of

Design	1	2	3	4
layer 1	14, 0.6	16, 1.0	15, 0.4	11, 0.8
layer 2	8, 0.8	13, 0.4	14, 1.8	9, 0.8
layer 3	6, 1.0	13, 0.6	11, 1.2	8, 0.4

Table 1: Output designs for 3-layer optimization problem from 0.1-10 GHz. Values shown are material numbers corresponding to those in [1] and thickness expressed in mm.

$\Gamma_{ga} = 0.996456 - j0.0841199$, which is excellent agreement.

Next, a two-layer Dallenbach scattering problem was set up with parameters: $t_1 = t_2 = 2$ mm, $\epsilon_{r1} = 10 + j0$, $\epsilon_{r2} = 50 + j0$, $\mu_{r1} = \mu_{r2} = 1.0 + j0$, and the frequency was again 1 GHz. The analytical result is $\Gamma_{exact} = 0.985 - j0.172$ and the compDesign uted result is $\Gamma_{ga} = 0.98502 - j0.172442$.

For a three-layer Dallenbach scattering problem, the parameters were: $t_1 = t_2 = t_3 = 2$ mm, $\epsilon_{r1} = \epsilon_{r3} = 10 + j0$, $\epsilon_{r2} = 15 + j0$, $\mu_{r1} = \mu_{r2} = \mu_{r3} = 1.0 + j0$, and the frequency was 1 GHz. The analytical result is $\Gamma_{exact} = 0.96489 - j0.2626495$ and the computed result is $\Gamma_{ga} = 0.964861 - j0.262761$. If we change material 2 to $\epsilon_{r2} = 50 + j0$, the results are then $\Gamma_{exact} = 0.96075 - j0.27742$ and $\Gamma_{ga} = 0.960713 - j0.277545$. These Dallenbach scattering problems therefore show that the cost function is being evaluated correctly for various types of input situations involving lossless dielectric materials.

To further calibrate the genetic algorithm code and the associated reflection coefficient calculations, several optimal designs presented in [1] were input to the genetic algorithm code to test the accuracy of the reflection coefficient calculations. The designs tested were the LF1, LF2, LF3 and LF4, and HF1, HF2, HF3 and HF4. These results disagreed with the graphical plots presented in [1]. A separate computer program written in FORTRAN was developed to provide the analytical reflection coefficient for a stratified media based upon the theory presented in this report for the parallel polarization. The LF1-4 and HF1-4 designs were input to this code, which provide the same results as the C++ genetic algorithm code. This independent verification indicates a possible problem with the published results. However, due to time considerations, the exact cause of this discrepancy was not investigated.

To demonstrate the optimization algorithm, two sample problems were executed. The first problem is a three-layer optimization problem over the frequency range 0.1-10 GHz, normal incidence, with a maximum individual layer thickness of 2 mm. An initial population of 8 chromosomes was used, which results in an iterative population of 4 chromosomes. Each chromosome was randomly initialized, and the genetic algorithm was applied to optimize the design. The output designs are shown in Table 1. The analytical reflection coefficient results for these designs are shown in Figure 4 at the end of this report. The graphs show that indeed the Design #1 shows the best *mean*, or average,

Design	1	2	3	4
layer 1	14, 1.0	3, 1.4	1, 0.4	6, 0.2
layer 2	12, 1.2	2, 1.0	16, 1.6	5, 1.8
layer 3	11, 0.6	16, 0.8	13, 1.0	4, 1.2
layer 4	9, 0.8	15, 0.4	13, 1.2	2, 1.4
layer 5	8, 0.2	13, 0.4	12, 0.6	1, 0.8

Table 2: Output designs for 5-layer optimization problem from 0.1-10 GHz.

reflection coefficient over the entire frequency band, and it was chosen as the optimal design.

The next sample problem was the same, except the number of layers was increased to 5. The results for this case are shown in Table 2. The reflection coefficient results for each design in this problem are graphed in Figure 5. Note that although Design 4 shows a large null in the reflection coefficient, it appears that Design 1 has the lowest *mean* reflection coefficient over the frequency band. The difference between the final mean cost functions between Designs 1 and 4 is approximately 8 dB.

A final, larger problem was run which included an initial population of 64 chromosomes, an iterative population of 32 chromosomes, 100 iterations, 5 layers, 10 incidence angles from $0 - 90°$, and the same frequency band of 0.1-10 GHz. The first two output design reflection coefficient results for normal incidence are shown in Figure 6. Note that Design 1 offers the lowest mean reflection coefficient over the frequency band of interest.

3 Conclusions

The goal of this project to create an optimization algorithm for optimization of radar absorbing material layers has been achieved. The algorithm is based upon a binary genetic algorithm, which optimizes a reflection coefficient cost function over a range of frequencies and incidence angles of interest. The code was verified against analytical solutions and attempts were made to verify the code with previously published results in this area. The algorithm was demonstrated on various optimization problems to illustrate its operation. Future work for this project would include extension to the perpendicular polarization, enhancements to the genetic algorithm via improvements to weight the cost function with a measure of layer thickness or overall thickness, further validation and verification efforts and further improvements in mating/crossover procedures and algorithm convergence studies.

References

[1] S. Ranjithan E. Michielssen, J. Sajer and R. Mittra, "Design of lightweight, broad-band microwave absorbers using genetic algorithms", *IEEE Trans. Antennas Prop.*, vol. 41, no. 6/7, pp. 1024–1031, June/July 1993.

[2] James R. Wait, *Electromagnetic Waves in Stratified Media*, IEEE/OUP, New York, 1996.

[3] R. L. Haupt and S. E. Haupt, *Practical Genetic Algorithms*, John Wiley & Sons, New York, 1998.
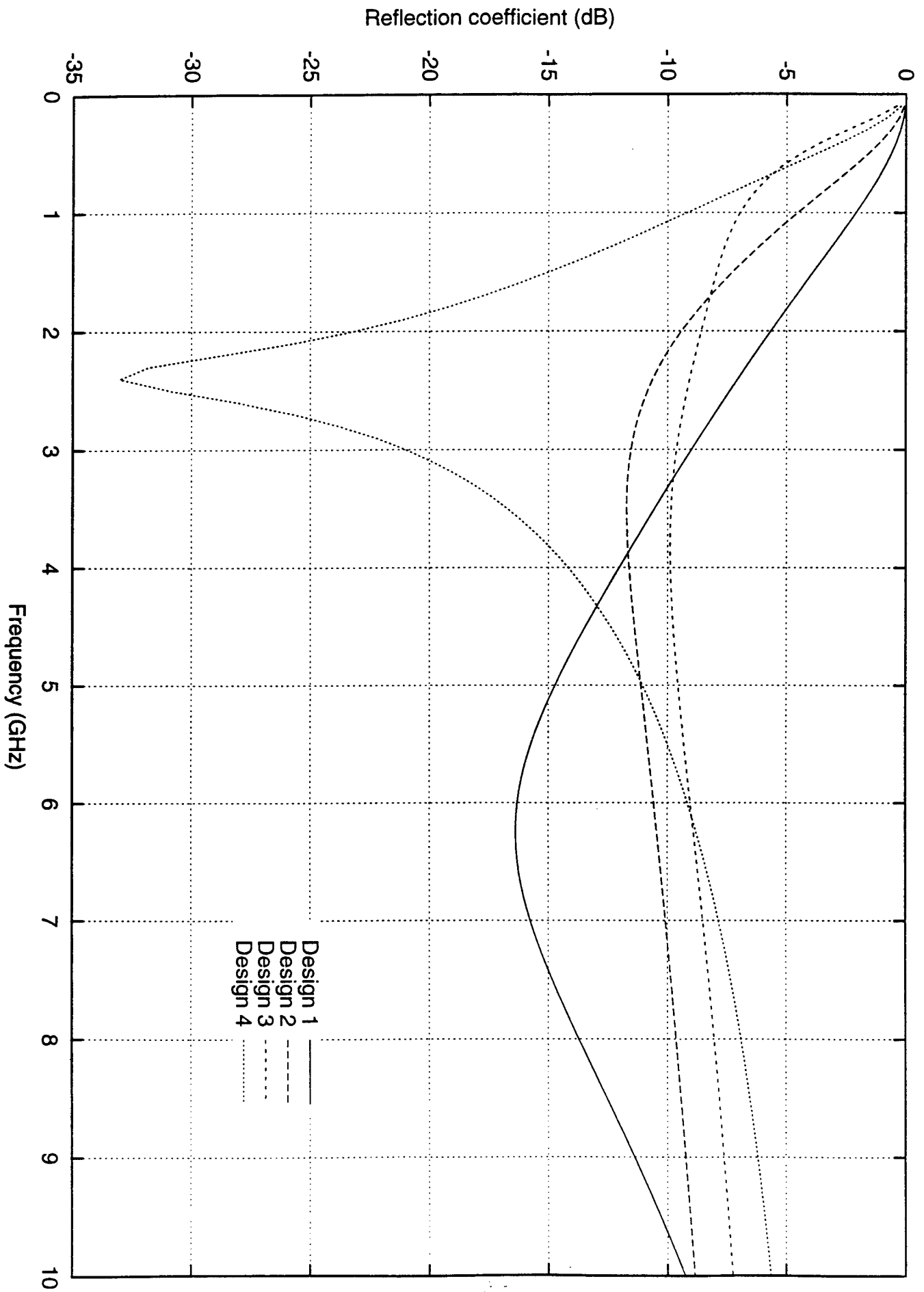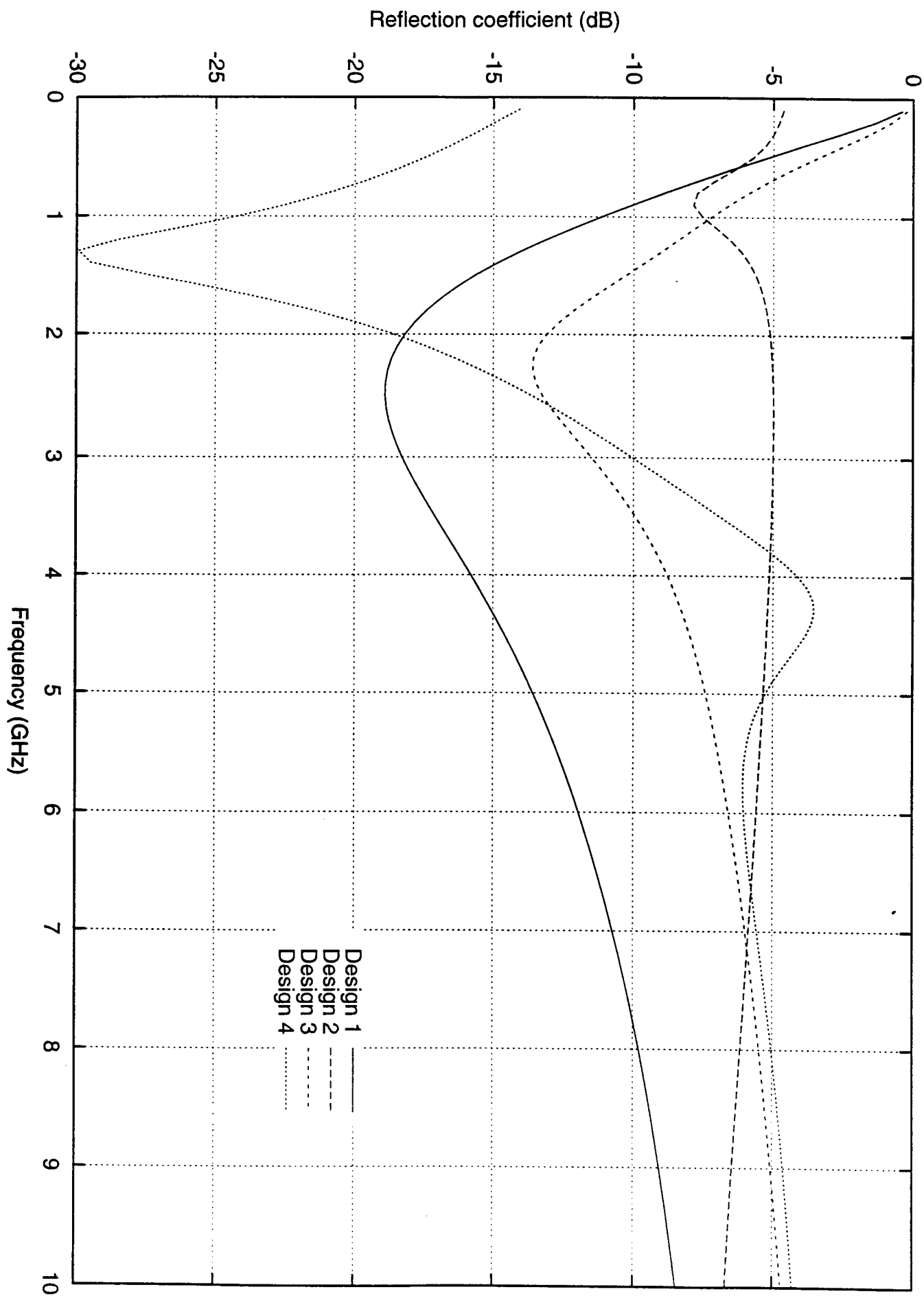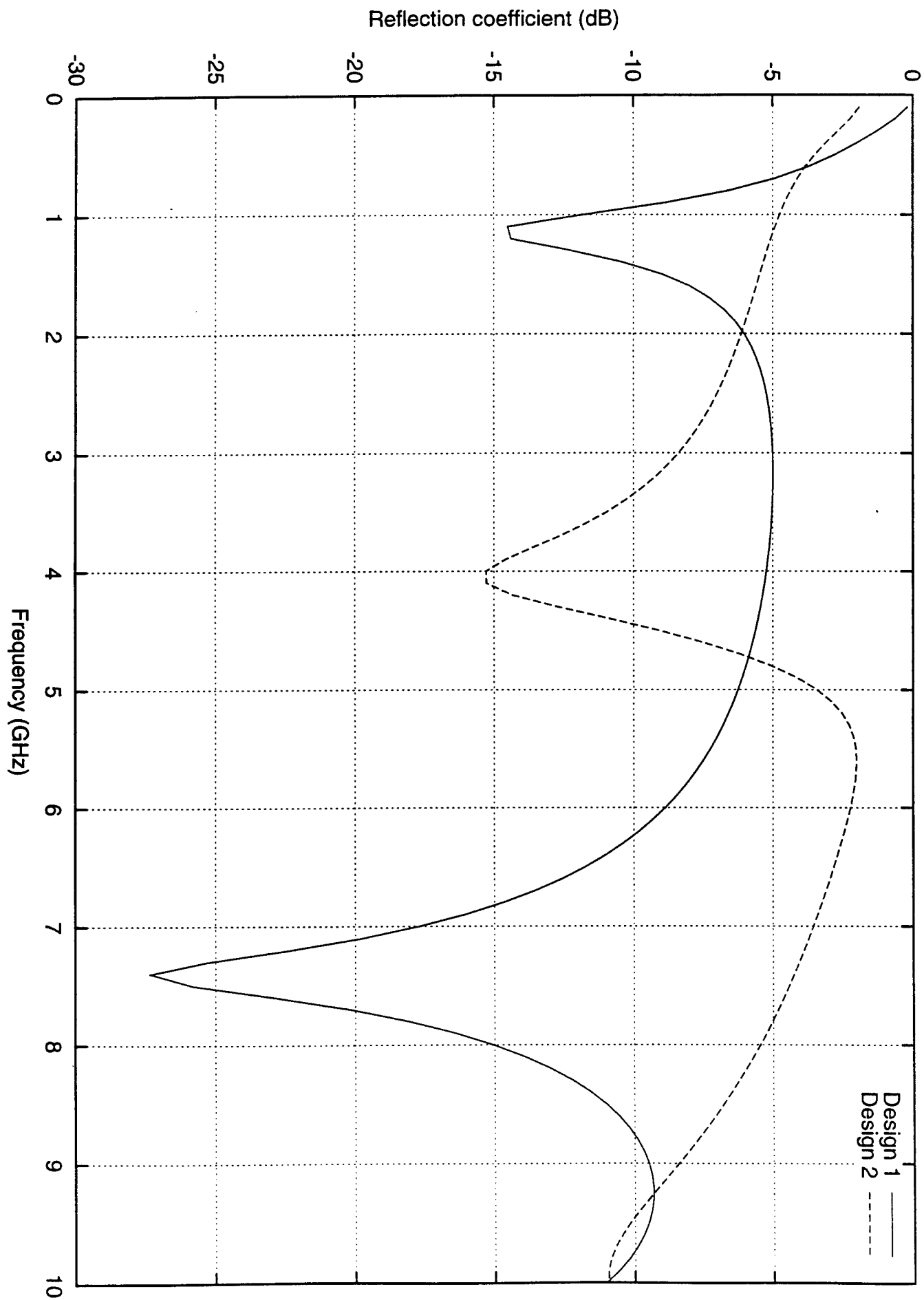
Figure 4

Figure 5

Figure 6

AFOSR Summer Research Extension Program
Final Report
Subcontract 98-0819

submitted by:
Raj Bhatnagar
Associate Professor
University of Cincinnati
Cincinnati, OH-45221-0030

Analysis of Intra-Class Variability and Synthetic Target Models for Use
in ATR

In Cooperation with: Air Force Research Laboratory
Wright Patterson Air Force Base
Dayton, OH

Setember 1999

1 Problem Description

In an Automatic Target Recognition system, vehicles are classified as belonging to either a target class or a confuser class (the set of non-target vehicles). Typically, this classification is achieved by selecting a member of the target class, and then creating a set of templates based on the images of the selected vehicle. The set of templates are used to represent the entire target class.

The classification decision is determined by the Mean Square Error (MSE) distance from an unknown image to the set of templates. A small distance indicates a close match to the templates, and therefore the unknown image is classified as belonging to the target class. Otherwise, the image is determined to belong to the confuser class.

One problem with this method arises when classifying certain classes of vehicles, such as the T72 class. Some members of the target class are classified correctly, while others are classified as confusers. The performance depends heavily on the vehicle of the target class that is chosen to create the templates. If there is a large degree of dissimilarity among the vehicles in the target class, then more than one set of templates is required to accurately represent the entire class. A solution is to design class models that together represent the entire target class. Such models are created by combining images from several different vehicles in the class.

One of the issues when building class models is determining the number of sets of templates to be used for a given target class. Another issue is selecting which images of the target class to use to create the sets of templates. This thesis presents methods for handling these issues and building class models.

2 A Pattern Recognition Problem

The goal of air-to-ground Automatic Target Recognition (ATR) systems is to detect potential targets in an environment with clutter in the background and then to determine whether the object in question is in-fact a desired target. An ATR system requires several steps to accomplish this task, and uses several components. An introduction to ATR systems can be found in [?]. This thesis has been placed in the context of ATR, where the sensor data are two-dimensional images from a synthetic-aperture radar (SAR).

The techniques that have been presented are aimed at discriminating between target vehicles and other confusing vehicles. The context of our work is the recognition problem when the target image has already been extracted from the environmental ground clutter surrounding it. The particular target class vehicles that were used for testing included a set of eleven T72 tanks. A complete description of the data set is given in appendix ??.

3 Template-Based Classifiers

The problem described in section 2 is essentially a two-class discrimination problem, where the vehicles belong to either a target-class or a confuser-class. A classifier must be designed to determine which class a given SAR image belongs to. One way of achieving this is to train the classifier with the images of a member of the target class. Test images from all the other vehicles are then classified based on how closely they match the classifier templates.

The SAR images of vehicles that are used for both training and testing are also known as *chips*. Each image chip is a two-dimensional, 128 by 128 gray-scale image, representing the vehicle at a specific aspect angle (0–359 degrees). The depression angle, with respect to the sensor collecting the SAR data, is 17 degrees for all the data used in this study.

The classifier is implemented as a set of templates, where each template represents a section of the vehicle used for training. Each template represents a five-degree section of the vehicle, and is created from the images of the vehicle included in the five degree angular window. Since each section is a five-degree window, seventy-two templates are needed to represent the entire three-hundred and sixty degrees of a training vehicle.

To build the individual templates, the image chips of the vehicle that fall within each five-degree window are collected. For example, the ten-degree template of a classifier is formed by collecting all image chips of the training vehicle representing aspect angles 7.5 degrees to 12.5 degrees (10±2.5). The image chips are aligned, and then averaged together to form the template for that section.

3.1 Classifying images

With template-based classifiers, the distance from a template to the unknown image is used for classification. The test image chip must be compared to each of the seventy-two templates, since the aspect angle of the test image is unknown. The distance to the closest matching template is used for classification. The program used to compute these template-chip distances is the "bright-region MSE classifier" (brmse), a program written by Ron Dilsavor of Sverdrup Technology. It implements Sandia's bright region MSE classifier design.

The performance of a classifier can be measured by how well it discriminates members of the target class from members of the confuser class. The target class used in this thesis is a set of T72 vehicles, and the confuser vehicles are all non-T72 vehicles. The template vehicle, or training vehicle, will be selected from the set of target vehicles.

Two types of distances are collected to test performance: Distances from the template to the target chips and the distances from the template to the confuser chips. Figure 1 shows how these two collections of distances are distributed, using T72 model number A32 templates as an example. The solid-line is a histogram representing the distribution of distances between image chips from the entire class of T72 vehicles and the closest matching A32 template. In contrast, the dashed-line histogram shows the distribution of chips from the set of confuser vehicles, again using each chip's distance to the closest matching A32 template. It is the overlapping portion of the histograms, in this case approximately .1 to 1, that makes the classification decision difficult. The more that these two distributions are separated, the easier the classification. The following section introduces a way to compare the relative performance of two or more templates, in terms of their ability to separate these distributions.

3.1.1 ROC curves

A standard way of measuring the performance of a template classifier is in the form of a Receiver Operating Characteristics curve, or a ROC curve. A ROC curve is a plot of the probability of detection (PD) versus the probability of false alarm (PFA). The PD is the probability of classifying a vehicle as a target, given that it is a target. The PFA is the probability of classifying a vehicle as a target, given that it is a confuser.
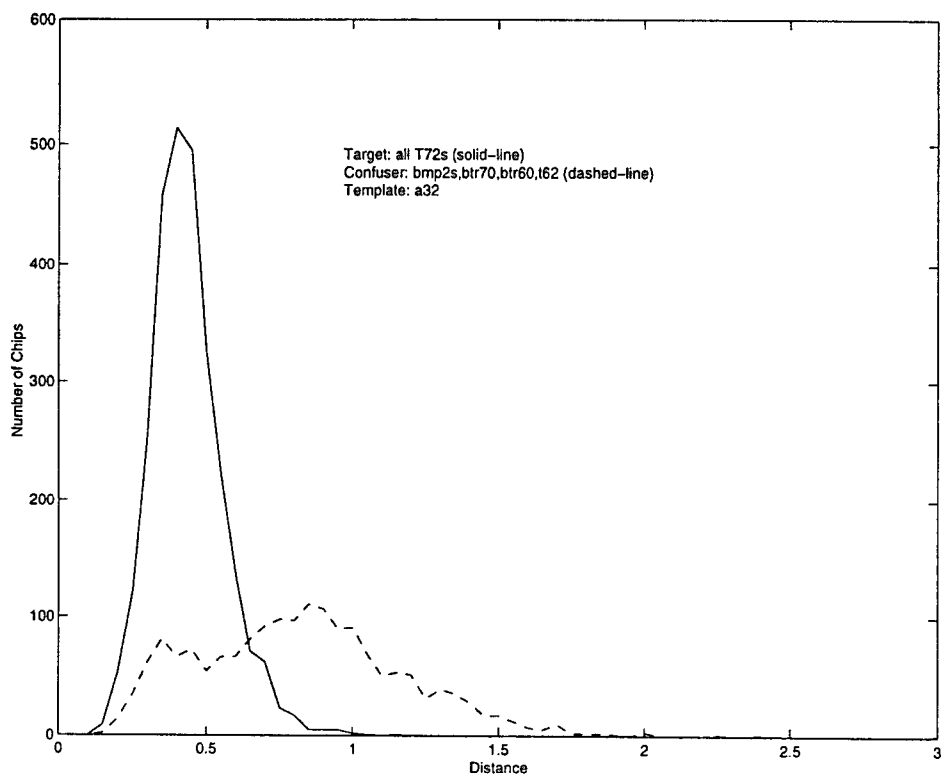
Figure 1: Distribution of template-to-chip distances
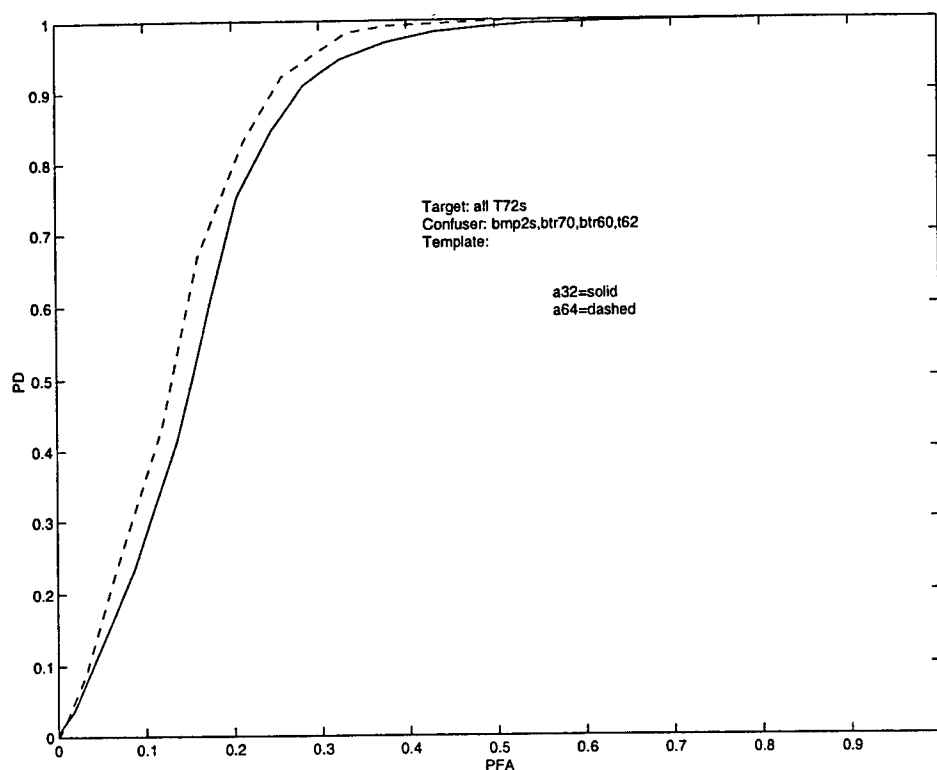
Figure 2: ROC curves

The ROC curve is generated using the two distributions of distances previously described, and demonstrated in Figure 1. The PD coordinate is determined by the number of chips in the target class with a distance less than some D, divided by the total number of target chips. As the distance D increases, more chips are included in the count, and the PD increases. These points are plotted against the PFA coordinate, which is the number of chips in the confuser class with a distance less than the D, divided by the total number of confuser chips. As more confuser chips are included in the count, by the increasing distance D, the PFA also increases.

Figure 2 demonstrates a comparison of two ROC curves. The target class is all eleven T72 vehicles, and the confuser class contains three BMP2 variants, the BTR70, the BTR60, and a T62. A detailed listing of the vehicles used in the comparisons can be found in appendix ??. The solid-line ROC curve shows the result of using T72 model number A32 templates as the classifier, and the dashed-line ROC curve shows the result of using T72 model A64 as the classifier. Figure 2 indicates that an A64 discriminates the class of T72s better than an A32 vehicle, since at any point the probability of detection is higher for a given probability of false alarm.

In all of the experiments presented in this thesis, the vehicle used for training is excluded from the target class. For example, with the A64 classifier in Figure 2, the distance from the A64 template to the A64 image chips was withheld when generating the ROC curve. The same was done with the A32 ROC curve and the A32 template-chip distance. Only the template distances to the other ten vehicles were compared to the confuser chip distances. This is done

so that a comparison does not reflect a bias when a vehicle used for training is the same vehicle the classifier is tested on.

4 Effects of Intra-Class Variability

The accuracy of the classification depends on the vehicle chosen for forming templates, since that is essentially the vehicle to which a test image is compared. The problem is that a class of vehicles may consist of members that are very different from the other members of the class. This variation within a class makes it difficult for a classifier to discriminate target class vehicles from confuser class vehicles.

The reason is that a classifier designed with any one member of the class may perform very well for some members and very poorly for other members. For example, a classifier designed with a T72 model number A62 performs well when tested with a model number A63, but not as well when tested with an A05. The vehicles that are most similar to the training data can be classified more accurately. If a class has few vehicles that are similar, or has a high degree of variability, then the classification performance for any unknown member of the class will be unpredictable.

The eleven T72 vehicles are a class with a large degree of variability. Figure 3 demonstrates this, with a comparison of several ROC curves. Each curve shows the performance of particular T72 templates when discriminating the class of T72 targets from the confuser vehicles. It is evident from the plot that the performance of the classifier varies significantly depending on which T72 is used to create the templates. For example, although model number A10 performs relatively well, model number A32 does poorly when discriminating targets from this large set of confusers. The class of T72 vehicles illustrates the problem of intra-class variability when designing template-based classifiers.

5 Goals of This Project

The objective of this research project was to design algorithms for constructing models of classes of vehicels and to determine the effectiveness of such models for Automatic Target Recognition (ATR) Systems. A class of vehicles, such as the T-72s, has large variation among all the members of this class. Due to this variability, an ATR system trained on one member of the class may not be able to recognize all the other members as belonging to its own class. It is desirable, therefore, to design models of vehicle classes that deemphasize the intra-class variabilities while retaining enough information to discriminate against non-members of the vehicle class.

Section 8 introduces a measure for determining the theoretical compactability of a class of vehicles into a smaller number of representative models. This measure can help in choosing an appropriate number of models, with the corresponding expected error, for a class of vehicles. Section 9 presents a simple clustering algorithm to select the vehicles to be merged together while forming templates for the combined representative models. The approach in Section 10 incorporates the aspect-angle dependent variability of the class in the model building process, and thus provides more flexibility to the selection of images for constructing class models. In Section 11, the experimental results are shown for classifiers designed usingboth the above
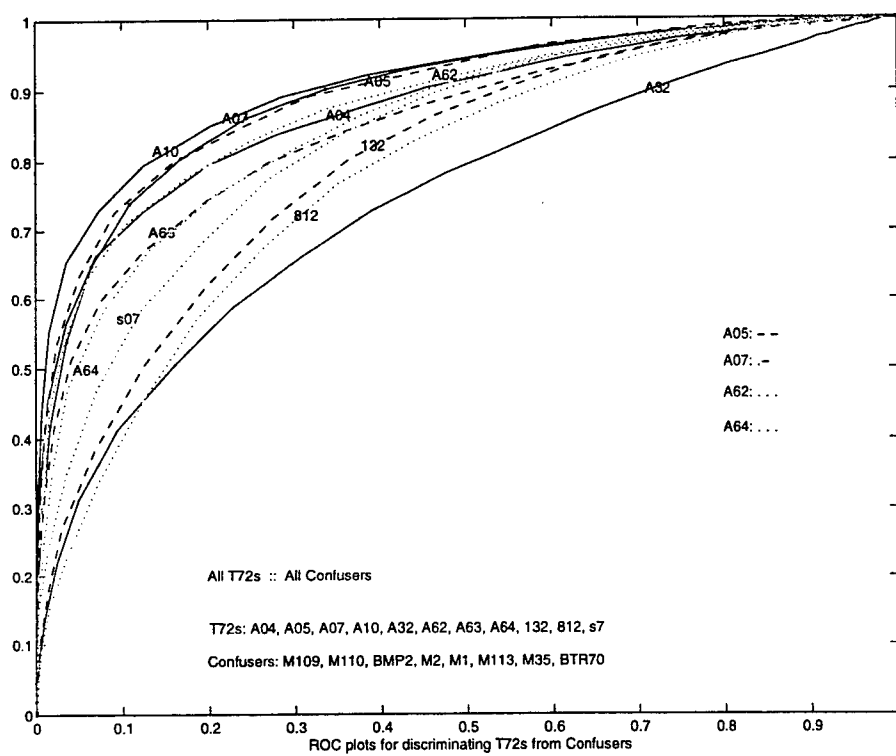
Figure 3: Intra-class variability of T72 vehicles

class models. We show significant improvement in performance by these class models.

6 Related Work

The difficulty of separating vehicles in the same class, especially the T72 class, is demonstrated and explored in [WPB97] using template-based classifiers. One approach taken was to train a classifier with templates that are created by averaging the templates of all vehicles in the target class. Improvement was shown for a class of three T72 variants. The other techniques developed in [WPB97] involved normalizing and thresholding the images before the templates were formed. In this thesis, the benefits of mixing templates is further examined on a much larger set of T72s, where the focus is on creatively selecting images of the different vehicles to be used in template formation. Also, in chapter ?? a thresholding scheme is shown which processes images after template formation.

In [SKW+98], a classifier is designed with models using MSTAR Predict, to compensate for various target poses and intra-class target variability. MSTAR Predict produces a simulated radar image from a CAD model, which can be used to classify images according to the vehicle modeled. The paper shows that these models are still susceptible to the problem of intra-class variability, which is "generally more difficult" than handling target pose in Predict models [SKW+98]. Results are mentioned with improved performance by including information about the intra-class variability of T72 models. This thesis complements their work be developing techniques to extract information about the T72 class variability for better classification. Also, our classifiers are designed using actual measured SAR images, and not predicted signatures. Although [SKW+98] demonstrates the use of the MSE metric with MSTAR Predict models, no steps are taken in this thesis to show how our results apply to these models.

A detailed study of intra-class variability is given in [BDMP98]. The performance of templates created from combining images is tested, using a few types of vehicles from the T72 class. The intention of this thesis is to extend the work on combining images, providing techniques to determine how many and which images to combine. [BDMP98] also performs tests using signature prediction with DEMACO's XPATCH system, modeling a specific T72 vehicle. The classifiers in this thesis were only trained on measured SAR data.

7 Class Models

The intra-class variability of target vehicles degrades the performance of a classifier. The approach taken in this thesis is to create class models, or a number of subgroup representative classifiers that together represent the entire class of T72s. This improves the overall performance when discriminating such classes against confusers. A class model is based on a set of data to train the classifier, that is meant to represent an abstract type of vehicle in the class. They are created using images from several different tanks in the class.

Using the T72 class as an example, one such class model may be designed to represent vehicles A04, A05, A07, and the A10, assuming these vehicles belong to the same abstract type of T72. If there are four subgroup models used to represent the entire class, then all four models would be required to adequately discriminate all target vehicles. This is considerably less computation than using each of the eleven individual vehicles to discriminate the targets

from confusers. Also, since information about each vehicle is incorporated into the models, they should perform better than any individual vehicle in the class.

7.1 Class models with templates

Since template-based classifiers are being used, a class model will be represented by a set of templates created using measured SAR images of vehicles. Each template will be formed by averaging a list of images from (potentially) several different vehicles. If this is done for each aspect angle, then the performance of the model can be tested in the same manner as templates that were created from a single vehicle.

Two important issues arise when designing such class models. The first issue is determining the number of models to be used for a given class. This will depend on the degree of variability inside the class. The second issue is deciding which particular images will be used to form the templates of a certain model.

8 Class Compactability

A first step in building class models is determining how many models should be used to represent a single class. This number will depend on the degree of variability within the class. A class with high variability would require more models to represent it than a class with less variability. For example, a class of vehicles may consist of many similar members, or it may have several very different members with only few being relatively similar to each other. Only similar vehicles could be represented accurately by the same model. A notion of compactability is introduced as a guide for determining the number of models to be used to represent the class.

8.1 A Measure of Compactability

Informally, the compactability of a class indicates the extent to which it can be represented by fewer number of models without losing information required to classify individual members of the class. If a class has a high degree of compactability, than the members of the class are relatively similar to each other. A class with lower compactability measure would indicate members that are relatively dissimilar.

The N known members of a class can be represented as N points in a distance space. A point has a location in this distance space defined by its distance from every other member in the class, therefore, resulting in an $N - 1$ dimensional space. As the number of dimensions used to represent the N members is decreased to $N - 2, N - 3, \dots, 1$, information will be lost and some error will be introduced. The dimension m, such that if N was reduced to $m - 1$, then the amount of information lost increases significantly, determines the compactability of the class.

For the class of T72 vehicles, each point in the space is a member or vehicle. The distance between any two points is the MSE distance between the images. As we represent the points of the space in fewer and fewer dimensions, we are assuming fewer types of vehicles required to represent the class. The structure of the class is shown by the distances between the vehicles. The compactability testing is a measure to determine how many of the vehicles are similar enough to be represented by the same class model.

Table 1: Table of vehicle distances

	132	812	S7	...	A64
132	0	.43839524	.40990450		.45779017
812	.43839524	0	.43240280		.44535691
S7	.40990450	.43240280	0		.43279583
⋮					⋮
A64	.45779017	.44535691	.43279583	...	0

8.2 Distance Metric

To find the distance between two vehicles, the average MSE distance between an image and the nearest template from the other vehicle is computed. This is repeated for all pairs of vehicles, so that a distance is obtained from each vehicle to every other vehicle in the class. Table 1 shows a sample table.

In the case of the class of T72s, there are eleven vehicles. The eleven vehicles can be treated as points and placed in a ten-dimensional space, preserving the pairwise distances discovered in the previous step.

The next step for computing the compactability is compressing the space. The points for the eleven T72 vehicles are compressed into successively smaller number of dimensions, preserving the pairwise MSE distances as much as possible. For example, the space will be reduced from ten to nine dimensions, then from ten to eight dimensions, continuing until finally ten dimensions is reduced to one dimension.

8.2.1 Algorithm for Compaction

The amount of error introduced by compressing the space has been measured in two ways. It has been measured in terms of the average MSE between the distances in the original space and the new distances in the compressed space, which is calculated during the mapping previously described. It has also been measured in terms of the reversed MSE distance pairs in the compressed space. The method used for calculating this second measure is shown in Figure 4.

This algorithm measures how many distance pairs had smaller MSE distances than other pairs in the original space but have larger distances in the compressed space. Starting with a distance D_i in the compressed space, its position or rank in the original space is found. Each distance D_j that is below D_i in the original space is tested to see if it is still below D_i in the compressed space. If the difference between the distances D_i and D_j is less than one percent, then the two distances are similar enough not to consider reversed. Otherwise, the relationship between the two points has not been preserved and the pairs are considered to be reversed.

As the number of dimensions decreased, the MSE of the new distances between points continued to increase. The rank order, as shown in Figure 5, did generally increase as the space was compressed, but not in every case.

The compactability results indicate that relatively more information about the structure of the class is being lost after the sixth and second dimension. At a low number of dimensions such as two, a large degree of error is introduced in terms of both MSE and reversed distance

Figure 4: Measuring reversed MSE distance pairs

Given a table T_a of MSE distance pairs in the original space, and a table T_b of MSE distance pairs in the compressed space,

- Sort T_a and T_b in non-decreasing order, creating $order_a$ and $order_b$ respectively.

- For each distance D_i in $order_b$:

    1. Find index k of D_i in $order_a$

    2. For every distance D_j in $order_a$ with an index greater than k:
        Find index l of D_j in $order_b$
        if l is less than i, then
            Compute the difference between D_i and D_j
            If the difference is greater than 1 percent, then
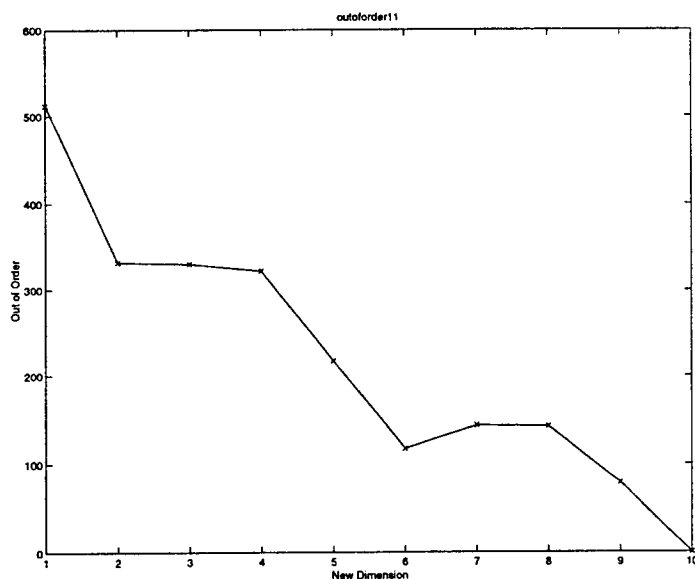   increment $out - of - order$ by 1

Figure 5: Impact of compressing space on rank order of distances

Figure 6: Clustering Algorithm

1. Create dissimilarity coefficient matrix

2. Sort rows of dissimilarity coefficient matrix in non-decreasing order by distance

3. Initialize all points as belonging to their own cluster (p clusters)

4. Scan sorted dissimilarity coefficient matrix, starting with the shortest distance.

5. If the two points with that distance belong to different clusters, then merge the two clusters they belong to.

6. While the number of clusters is greater than k, repeat to step 4

pairs. This is evidence for using six class models to represent the class of T72 vehicles.

9 Clustering

Compactability gives some insight into how many models can be used to represent a class. Once the number of models that will be used to represent a class is chosen, it is another matter to decide exactly which vehicles actually form the different types to be modeled.

One approach is to have each class model represent an exclusive subset of the vehicles in the class. In this way, the four models together could discriminate the class of vehicles. An unknown image would only need to be compared to each of the four models to decide if it belonged to the target class. Clustering can be used to decide which subset of vehicles each class model will represent.

To find a subset that only contains one type of vehicle, the vehicles should be grouped into sets of similar vehicles. The MSE distance between vehicle images can be used as a measure of similarity. If there are eleven vehicles, then each vehicle can be represented by a point in a ten dimensional space. The distance between any two points is the MSE distance between the corresponding vehicle images. The subsets of vehicles can be found by an exclusive classification of these points. The classification is exclusive to achieve disjoint partitions of the space (vehicles). This can be accomplished through clustering, where the number of clusters to be found will be the number of models to be used for that class.

9.1 Clustering Algorithm

The vehicles will be clustered according to their mutual MSE distances (the same distances that were used during compactability testing). Since the objective is to find points that are nearest to each other, a nearest-neighbor algorithm applies naturally. The steps taken in the algorithm are shown in Figure 6, where p is the number of points in the space and k is the number of clusters to be found.
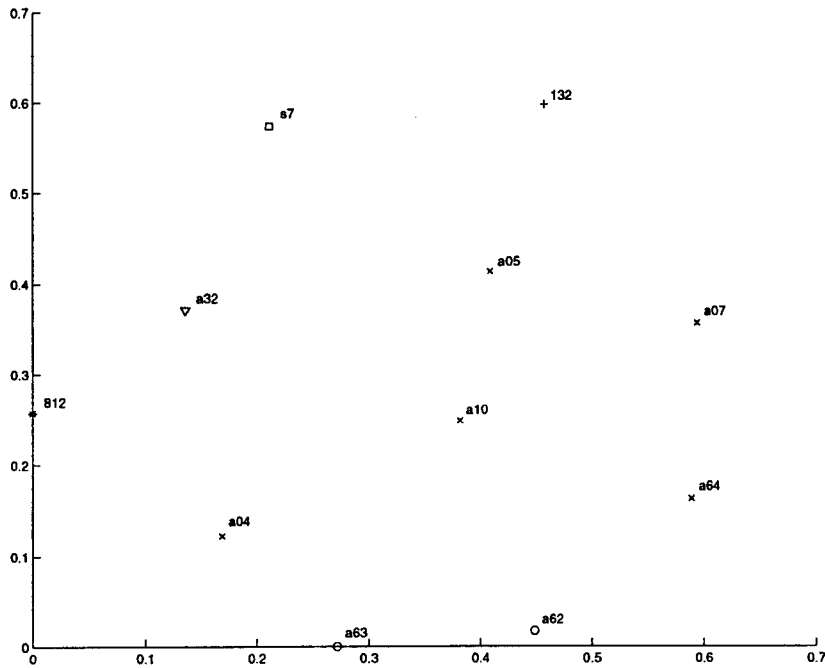
Figure 7: Clusters found for eleven vehicles

9.2 Vehicle Clusters

The algorithm was applied to the T72 class of vehicles, represented by eleven points in an eleven-dimensional space. The number of clusters to be found was set to six, as was indicated by the compactability results. Figure 7 shows the resulting clusters. The axes represent the coordinates of the points after being projected into a two-dimensional space.

One large cluster was found, consisting of the A04, A05, A07, A10, and A64 vehicles (labeled as an 'x' cluster). Another smaller cluster of two vehicles, A62 and A63, was found (labeled as an 'o' cluster). The other four T72 vehicles remained members of their own clusters, and were never merged with other vehicle clusters.

This implies that a large portion of the T72 class (five of the eleven) represent the same type of T72, and a second type of T72 is made up of the smaller 'o' cluster. The other vehicles have no vehicle in the class similar enough to be considered as representing the same type. In this way, those four vehicles (132, 812, S7, A32) are their own type of T72.

To build the class models, the same approach that is taken when forming templates for a classifier created from a single vehicle is taken when assembling templates for a model with images of more than one vehicle. The class model will consist of seventy-two templates covering five-degrees each. For each template, all the image chips falling within the degree range are collected from all the vehicles in the cluster, instead of from a single vehicle.

10   Aspect-Dependent Clustering

The previous approach clustered the images in a vehicle space, meaning they only allowed images to be considered as part of a particular vehicle. If this restriction is relaxed, we could look at the similarities in the vehicles across aspect angles. In this way, the image chips would belong to a degree-range group instead of a T72 group. For example, the image chips in the zero degree range (±2.5) from all different T72s can be clustered to form groups of T72 images that are similar from this aspect angle. If models are built based on the angle specific clusters, then it may be a more accurate representation of the class.

Another restriction of the vehicle-space clustering is using a fixed number of clusters when creating the templates used in a model. This technique does not reflect the fact that there is more variability among the images of certain aspect angles than others. For example, the front of a tank contains much more detail than the rear of a tank, and such detail can be exploited by a well designed classifier.

Clustering will be used to find similar image chips in the chip-space. As before, the clustering will determine which images are used to construct each class model. When clustering in the vehicle-space, the distances between vehicles had already been defined for use in comparisons. The corresponding comparison is not defined in the chip-space. In order to calculate the clusters of image chips, a notion of mutual distances between all image chips needs to be formulated. The MSE distance for each chip to a common template was used.

10.1   Clustering algorithm

Figure 8 describes the algorithm used to cluster the images using a distance between chips in the same degree range. Step one initializes a table of image chips versus their distance to each T72. There are as many columns as there are vehicles in the class, and as many rows as there are chips in the degree range. The table is then iteratively searched to find the minimum values. If two minimum values are in the same column, that indicates the chips of those rows are nearest to the same vehicle, so their clusters can be merged. If two minimum values are found in the same row, then the chip of that row has already selected a vehicle that it is closest to, so the chip should not be clustered based on the second minimum value. The number of clusters is continually reduced by merging, until a desired number of clusters is found. With the T72 data, the number of clusters used was four. After all chips are clustered in the first degree range, a new table is initialized for the next degree range and four new clusters are formed from the new list of chips. This continues until all aspect angles (spanning five degrees) are clustered, each containing four groups.

10.2   Variable Number of Clusters

The next step is to accommodate for a variable number of clusters for each degree range. There will need to be some restrictions placed on the data, such as the minimum size of a cluster and the maximum number of clusters. Also, the number of clusters should reflect the similarity of the images at each aspect angle. These requirements were met by running the above clustering algorithm in two passes.

During the first pass the chips are clustered into the maximum number of clusters that is to be allowed. For this particular data, The image chips were clustered into four groups. After the

Figure 8: Clustering Algorithm

1. A given chip is compared to eleven templates - one from each of the T72s. This is repeated for all chips in the aspect angle window. Each chip is a row of a table, where the columns are the templates.

2. Find the minimum value in this table and mark it.

3. Find the next minimum value, m_1, in this table.

4. • If there is a marked value in the same row as m_1, do nothing.

   • Else if there is a marked value in the same column as m_1, then cluster those chips (rows) together and mark m_1.

   • Else just mark m_1.

5. Repeat to step 3, until all chips are clustered.

6. Repeat process (from step one) for all seventy-two degree ranges.

first pass, the cluster information can be processed. The information that will be used is the number of chips that belong to each of the clusters found. For each cluster with less than some threshold number of chips, a $Reduce - by$ value will be incremented. In this case a threshold of three was used, so that clusters with two or less chips would increment $Reduce - by$. After all the clusters have been inspected, the $Reduce - by$ total will be subtracted from the maximum number of clusters used in the first pass. This calculated value will be the new number of clusters to be found.

The second pass will start from scratch, re-clustering the entire set of image chips with the same algorithm as before. But during this pass, the number of clusters to be found will have been adjusted by the results of the first pass.

These two passes will be applied to each of the seventy-two angle ranges, achieving the variable cluster numbers at each range. Figure 9 is a plot of the clusters found for the T72 data set. There was more variability among chip images at the points where more clusters were required, such as around the zero degree range.

10.3   Creating Models

The class models are created by building templates for each degree range, as was done before. There will be as few as one template for a degree range, or as many as the maximum number of clusters allowed, depending on the individual aspect angle variability. Each template will represent a sub-class of the vehicles in the class, at a specific aspect angle.

The list of chips that will be merged to form a template at a certain degree will be precisely the list of chips belonging to the cluster for that degree. The same program that was used before to create templates is applied to these new lists of chips.
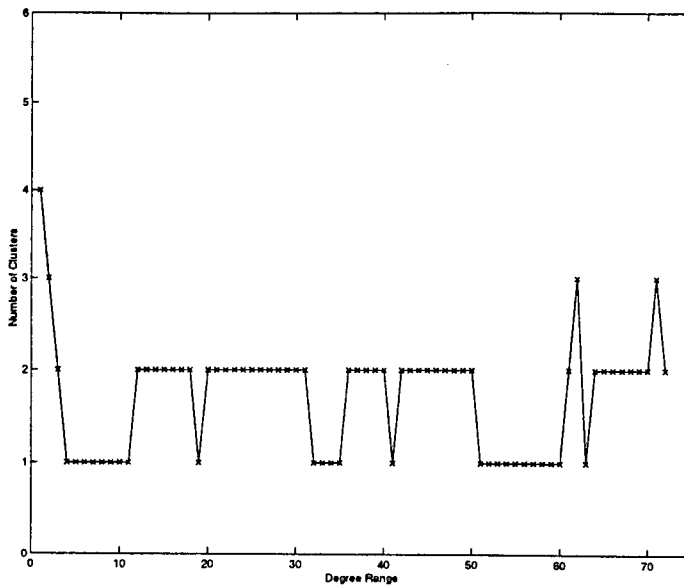
Figure 9: Variability of each aspect angle

11 Performance of Class Models

11.1 First clustering approach

This section presents the results of clustering in the vehicle-space. The performance of a class model designed using vehicles from a cluster is compared to the performance of classifiers trained on the individual vehicles of that cluster.

The ROC curves in Figure 10 compare classifier performance using six confuser vehicles. *cluster*1 is the class model created using vehicles from the large 'x' cluster of Figure 7. The other five templates demonstrate the performance of classifiers trained with the five vehicles of that cluster. The target class used in this experiment is also the five T72 cluster vehicles. The figure shows that the class model performed better than any individual vehicle of its cluster. Another experiment tested the performance of a class model designed using vehicles from the smaller cluster (A62 and A63). As before, the class model classifier out-performed classifiers trained on the individual members of the class.

11.2 Second clustering approach

This section presents the experimental results of a class model designed when clustering in the chip-space. The *mix* template is the class model classifier. Figure 11 shows ROC curves with a target class of all T72s and the entire confuser class. As the figure demonstrates, the class model performs better than any individual training vehicle, especially at the upper part of the curve where the probability of detection is higher.

The ROC curves in Figure 12 compare the two class model approaches, *cluster*1 being the vehicle-space model and *mixed* begin the chip-space model. The test is comparing with targets *cluster*1 was designed to classify, which are named again in the figure. The results show that *cluster*1 performs better, although not by a large margin.

3-16
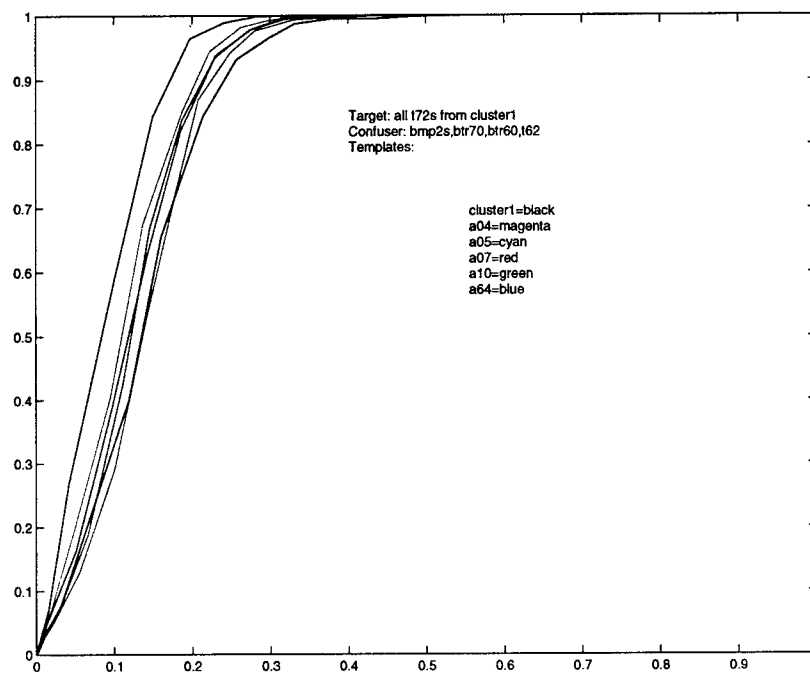
Figure 10: Class model from the vehicle-space (*cluster*1)
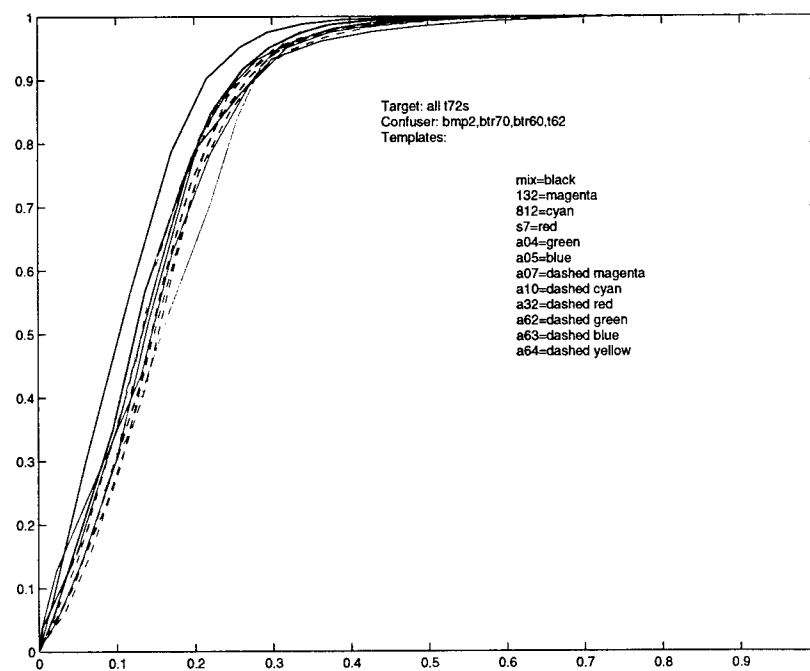
Figure 11: Class model from the chip-space (*mix*)
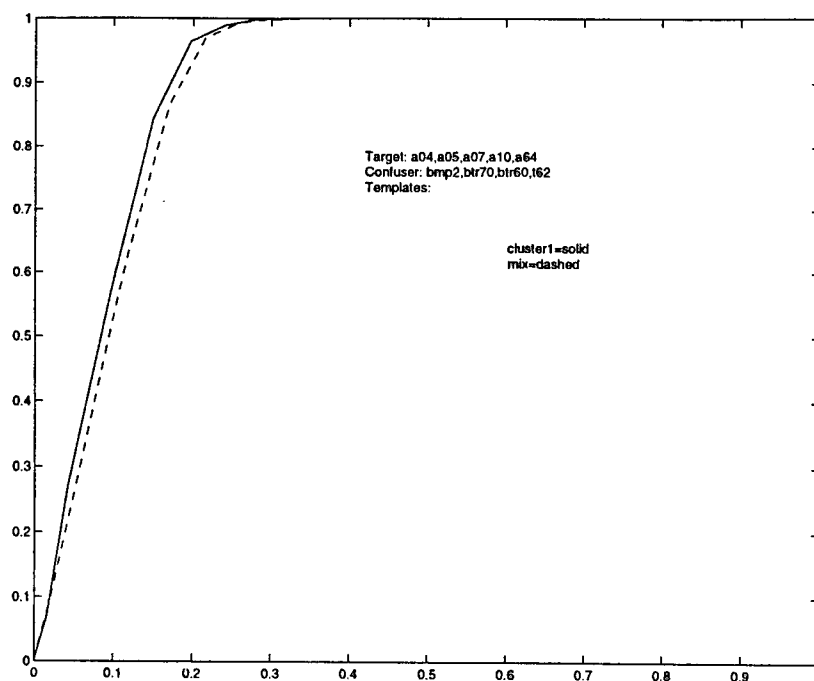
Figure 12: Chip-space model (*mix*) versus vehicle-space model (*cluster*1)

References

[BDMP98] Raj Bhatnagar, Ron Dilsavor, Mark Minardi, and Dax Pitts. Intra-class variability in atr systems. In Edmund G. Zelnio, editor, *Algorithms for Synthetic Aperture Radar Imagery V*, volume 3370, pages 383–395. Proceedings of SPIE, 1998.

[SKW+98] Stephen A. Stanhope, Eric Keydel, Wayne Williams, Vasik Rajlich, and Russ Sieron. The use of the mean square error matching metric in a model-based automatic recognition system. In Edmund G. Zelnio, editor, *Algorithms for Synthetic Aperture Radar Imagery V*, volume 3370, pages 360–367. Proceedings of SPIE, 1998.

[WPB97] Steve Worrell, Sharon Parker, and Mike Bryant. Class separability assessments and mse algorithm robustness. In Edmund G. Zelnio, editor, *Algorithms for Synthetic Aperture Radar Imagery IV*, volume 3070, pages 294–304. Proceedings of SPIE, 1997.

VALIDATION OF A LARGE EDDY SIMULATION CODE AND DEVELOPMENT OF COMMUTING FILTERS

G. A. Blaisdell
Associate Professor
School of Aeronautics and Astronautics

Purdue University
1282 Grissom Hall
West Lafayette, IN 47907-1282

Final Report for:
Summer Research Extension Program
Air Force Research Laboratory

July 1, 1999

VALIDATION OF A LARGE EDDY SIMULATION CODE
AND DEVELOPMENT OF COMMUTING FILTERS

G. A. Blaisdell
Associate Professor
School of Aeronautics and Astronautics
Purdue University

Abstract

A large-eddy simulation (LES) code developed at the Air Force Research Lab is applied to the problem of decaying compressible isotropic turbulence in order to help validate it. The code uses a high-order compact finite difference scheme coupled with a high-order compact spatial filter. Comparisons are made with previous direct numerical simulation (DNS) and LES results from a Fourier spectral code. For the DNS test case the finite difference code is found to give excellent agreement with the results of the spectral code. LES of a low Reynolds number case show that the finite difference code run on a 48^3 grid gives results similar to those of the spectral code run on a 32^3 grid. The results from the finite difference code run on a 32^3 grid for both the low Reynolds number test case and a high Reynolds number test case show some differences compared to the results of the spectral code. It is believed that the discrepancy is due to the solution filtering procedure in the finite difference code, although further simulations are needed to test this hypothesis. The issue of the commutation of filtering and differentiation is considered. The filters recently proposed by Vasilyev are found to produce commutation errors that are less than the truncation error of the finite difference scheme used. Several suggestions are made for ways to complete and extend the current work. Extensions include a detailed examination of the effect of solution filtering and its relationship to filtering as it is viewed in classical LES.

VALIDATION OF A LARGE EDDY SIMULATION CODE AND DEVELOPMENT OF COMMUTING FILTERS

G. A. Blaisdell

1   Introduction

Most flows of engineering interest are turbulent, and turbulence can have a large effect on the performance of engineering systems. For instance, the drag on a body is generally higher for turbulent flow, turbulence greatly increases heat transfer rates, and turbulent mixing is important to combustion systems. Numerical simulation of turbulent flows has become a valuable tool for gaining insight into the complicated flow physics of turbulence and it has provided data for evaluation of engineering turbulence models.

Direct numerical simulation (DNS) of turbulent flows involves solving for the time dependent motion of all the relevant length scales within a turbulent flow. Because the range of length scales increases with Reynolds number, it becomes prohibitively expensive in terms of computer resources to compute DNS of high Reynolds number flows. A method for overcoming this limitation is to directly solve for the large scale motions and to model the effect of the small scale motions. This approach is called large eddy simulation (LES).

In LES the small scales are averaged out using a spatial filter. As a result of the averaging procedure some information is lost which must be replaced with a subgrid-scale (SGS) model. The SGS model provides a closure for the subgrid-scale stresses, much the same as the way an engineering turbulence model provides information on the Reynolds stresses. However, it is believed that the large scales vary widely from one flow to another while the small scales are more universal. Therefore, in LES one can use a simple SGS model and, since the large scales are solved for directly, the results of an LES should be more robust than those of an engineering turbulence model.

In addition to its use as a research tool, large eddy simulation holds the promise of becoming a valuable engineering tool in analyzing problems that are too difficult for conventional engineering turbulence models to predict accurately. An example of such a problem is the inherently unsteady flow of a turbulent vortex interacting with a flexible structure, such as occurs on fighter aircraft at high angles of attack when the forebody vortices impact the tails. The unsteady loading on the tail structures causes premature cracking and is of serious concern to the Air Force. Another example in which LES is needed is the computation of turbulent jets and boundary layers to provide time dependent information for computational aeroacoustics calculations. An additional example is the use of LES in computing turbulent mixing in combustion problems, such as for fuel injectors in scramjet engines.

The primary objective of the current research is to help validate a large eddy simulation code developed by Drs. Miguel Visbal, Datta Gaitonde, and Donald Rizzetta of the Air Force Research Lab. The LES code uses high-order compact finite difference schemes based on the work of Lele[7] and refined by Gaitonde et al.[4] and Visbal and Gaitonde[15]. This code is written for compressible flow in generalized coordinates. It needs to be validated and tested on a variety of flows. A series of test cases is outlined in the report for the 1997 AFOSR Summer Research Program by Blaisdell[2]. The test case considered here is that of decaying compressible isotropic turbulence.

The LES code uses currently available LES technology. However, as pointed out in [2], there are several unresolved issues regarding the formulation of the LES equations and the SGS models. Resolution of these issues is important to the future development of LES. The second objective of this research project addresses

the issue of the commutation of filters with derivatives. Recent work is examined for filters that commute with an error that is less than the truncation error of the differencing scheme used for the LES calculations.

The purpose of this report is to summarize the progress made during the course of this project on these two objectives. The issues involved are not fully resolved at this point. Therefore, the report concludes with suggestions for future work.

2 LES Code Validation: Compressible Isotropic Turbulence

The simplest turbulent flow is decaying isotropic turbulence. LES of compressible isotropic turbulence were used by Moin *et al.* [8] to investigate their compressible formulation of the dynamic SGS model. Similar simulations were done by Spyropoulos & Blaisdell[12] and Spyropoulos[11] to more thoroughly investigate the ability of the dynamic model to capture compressibility effects and to examine several other issues in applying the dynamic SGS model.

The above LES and previous DNS of compressible homogeneous turbulence all used initial conditions which created artificial acoustic waves. Ristorcelli & Blaisdell[9] have devised a method of producing initial conditions which are more consistent and which do not produce strong acoustic waves. Such initial conditions should be used in any future studies of compressible homogeneous turbulence. However, since our purpose here is to validate the LES code, initial conditions are chosen to match those of previous simulations.

The governing equations and numerical method used in the code are described in Rizzetta et al.[10]. The LES code is called FDL3DI.LES. A detailed description of a version of the code which does not include an SGS model is given in Gaitonde and Visbal[5]. As mentioned above the code uses high-order compact finite difference schemes developed in [7, 4, 15]. It also uses a compact filter [7, 4, 15] in order to stabilize the simulations. The filter is applied directly to the solution at the end of each time step and, therefore, in the discussion below the procedure is referred to as solution filtering. It is believed that the filtering has an impact on the LES results, as discussed below. Some of the results for isotropic turbulence are included in Rizzetta et al., along with test results for channel flow and flow over a cylinder. Here more complete results are given for the isotropic test cases.

2.1 Description of the Simulations

The simulations considered here are compared to two cases from Spyropoulos & Blaisdell[12], Case 6 and Case CBC. Case 6 is a relatively low Reynolds number case with significant compressible fluctuations, while Case CBC is meant to reproduce the nearly incompressible experiment of Compte-Bellot and Corrsin[3].

The governing equations are nondimensionalized using a reference length, l_r, chosen so that the domain size has a length 2π on each side. The length scale of the turbulence is then defined by specifying the initial three-dimensional spectra. For Case 6, the initial spectra for the velocity, density and temperature fluctuations are of the form

$$E(\kappa) \propto \kappa^4 \exp\left[-2(\kappa/\kappa_p)^2\right] \tag{1}$$

where κ is the magnitude of the wavenumber vector, and the wavenumber of the peak of the spectrum κ_p is set at 4. The proportionality constant in the spectra is adjusted to set the r.m.s. level of the velocity, density and temperature. The r.m.s. level of the velocity is specified so that the initial turbulent Mach number $M_t = 0.4$, where M_t is defined as the ratio of the r.m.s. magnitude of the fluctuating velocity to the mean speed of sound. The reference velocity, u_r, in the nondimensionalization of the governing equations is chosen to be the initial mean speed of sound, so the reference Mach number is $M_r = 1$. In addition to the

turbulent Mach number, the velocity field is parameterized by the fraction of energy in the dilatational part of the velocity, χ (see [12] for details). In the current simulations the initial value is $\chi = 0.2$. The initial scaled r.m.s. density is $\langle \rho'^2 \rangle / \langle \rho \rangle^2 = 0.032$, and the initial scaled r.m.s. temperature is $\langle T'^2 \rangle / \langle T \rangle^2 = 0.005$, where the angled brackets, $\langle \ \rangle$, indicate a volume average over the computational domain. The reference Reynolds number, Re, is set to 536.9, which corresponds to a turbulent Reynolds number $Re_t = 2157$, where $Re_t = q^4/(\varepsilon \nu)$ and $\langle \rho \rangle q^2 = 2k = \langle \rho u_i'' u_i'' \rangle$ is twice the turbulent kinetic energy (k or TKE), ε is the TKE dissipation rate, and ν is the kinematic viscosity.

For the CBC case the initial velocity spectrum was chosen to match the experimental measurements of Compte-Bellot and Corrsin[3]. The initial turbulent Mach number was set to $M_t = 0.3$, the initial density and dilatational velocity fluctuations were zero. The pressure field was found by solving the Poisson equation for incompressible flow. The computational Reynolds number was $Re = 1588$, and the initial turbulent Reynolds numbers was $Re_t = 3166$.

Periodic boundary conditions are used for the isotropic simulations. Within the finite difference code, FDL3DI, periodic boundary conditions are implemented using a five-point overlap in each direction. In discussing grid sizes below, the five-point overlap region is not included in order to facilitate comparison with simulations using a spectral code which does not use an overlap region.

Several LES runs were made of the two cases using the current code, FDL3DI. A description of the current runs is presented in Table 1, along with a description of previous runs which are used for comparison. The run name is used within this report to distinguish the simulations. The case name refers to one of the two cases, Case 6 or CBC from [12]. Under "Code" spectral refers to the Fourier spectral code used by Spyropoulos and Blaisdell[12], while FDL3DI refers to the current LES code. The spectral code used the dynamic SGS model; however, FDL3DI has the option of using either the dynamic model or the Smagorinsky model. The main difference is that with the Smagorinsky model the model coefficients are constant, while for the dynamic model they are determined dynamically from the simulation data.

For the Smagorinsky model FDL3DI also includes the option of using different types of finite difference schemes to compute the strain rate used in the eddy viscosity. It is emphasized here that the lower order schemes were only used to compute the eddy viscosity and that all other derivatives were computed using the compact finite difference schemes presented in Gaitonde and Visbal[5]. All the possible finite difference scheme options were exercised for the LES presented in Table 1; however, only minor differences were found using the various schemes and, therefore, only the runs using the more accurate compact schemes are discussed in this report.

2.2 Simulation Results: Time Histories

Simulation I2 is used to validate FDL3DI as a DNS code. It is compared to results from the DNS I1, which used the previously validated spectral code of [12]. The evolution of the turbulent kinetic energy is shown in Figure 1. Here time is normalized as t/τ_0, where $\tau_0 = k/\varepsilon$ is the initial value of the eddy turnover time. Excellent agreement is found. The evolution of the density variance is given in Figure 2. Again, the agreement between the two codes is excellent. These results give confidence that FDL3DI is solving the governing equations correctly, as a DNS code.

Results from the 32^3 LES runs corresponding to Case 6, I3, I4 and I5, are shown in Figures 3 and 4 for the turbulent kinetic energy and density variance respectively. The results are also compared to the filtered DNS results from run I2. Here the DNS results have been filtered from a 128^3 grid onto a 32^3 grid. The turbulent kinetic energy for all of the LES simulations is below that of the filtered DNS. The LES results from the

spectral code are closer to the DNS data than the two finite difference code results. As is shown below by looking at spectra, the reduced TKE in the finite difference LES occurs because energy is diminished in the higher wavenumber range of the spectrum. It is believed that this is due to the explicitly applied filter, which removes energy from the higher wavenumbers in the finite difference calculation, although the effect of the filtering cannot be determined for certain without performing further simulations that do not use filtering. This point is discussed further throughout the report. The results for the density variance shown in Figure 4 show trends that are generally similar to those seen for the TKE.

Figures 5 and 6 show the TKE and density variance for the 48^3 LES corresponding to Case 6, I6 and I7. Note that the filtered DNS data is the same as that shown in Figures 3 and 4, because the DNS data was only projected onto a 32^3 grid and not onto a 48^3 grid. However, because the DNS is very well resolved, there is not a significant difference between the filtered DNS data and the unfiltered DNS data. Therefore, the available DNS data can still be used to give a meaningful comparison with the 48^3 LES results. Also, note that only 32^3 results are available from the spectral code. The 48^3 LES results for TKE lie closer to the DNS data than the 32^3 spectral code results, and they are much closer to the DNS data than the 32^3 finite difference code results. It seems that a finer grid is needed for the finite difference code in order to produce results as good as the spectral code. Again, the trends for the density variance are similar to those for the TKE.

With the dynamic SGS model the model coefficients C, C_I, and Pr_t are computed from the simulation rather than being specified *a priori*. A comparison of the coefficients is presented for the Case 6 simulations in Figures 7-9. The model coefficient used in the eddy viscosity, C, is shown in Figure 7. The spectral code gives values that are similar to the constant value of 0.00846 used in with the Smagorinsky model. The 32^3 finite difference simulation produces values of C that are greater than those of the spectral code, while the 48^3 simulation gives values that are closer to those of the spectral code. (Note that due to a post-processing error the values of C for the 48^3 finite difference LES presented in Figure 3 of [10] are too small by a factor of $(48/32)^2 = 2.25$. The agreement with the results of the spectral code is actually quite good.)

The evolution of C_I, which is used in the modeling of the trace of the normal SGS stresses, is given in Figure 8. However, since the finite difference code predicted negative values of C_I, the code set C_I to be zero. It is not known why the finite difference code produced negative values of C_I while the spectral code produced positive values. For the CBC case discussed below, the finite difference code produces positive values.

The last of the model coefficients is the turbulent Prandtl number, Pr_t. This is plotted in Figure 9. Both of the finite difference LES produce values of Pr_t that are similar to those from the spectral code, with the 48^3 results being closer than the 32^3 results. All of the dynamic model calculations are well below the value of 0.9 used in the Smagorinsky model.

The CBC case is at higher Reynolds numbers than Case 6. The history of the TKE and the density variance is shown in Figures 10 and 11 respectively. The finite difference LES results are compared with the spectral LES results. The spectral LES were compared previously with the experimentally measured spectra in Spyropoulos and Blaisdell[12], but the experimental data is not included here. DNS data for the CBC case has not been published, although a simulation has been done using a 512^3 grid (Alan Wray, NASA Ames, private communication). There is not a significant difference between the two finite difference LES simulations, indicating that differences in the SGS model do not have a large impact for these runs. The times corresponding to where experimental measurements were taken are $t/\tau_0 = 0.2870$ and 0.6612. At these times there is not a large difference in the values of the TKE among the three LES simulations. However,

there is a significant difference in the history of the TKE at early times between the spectral LES and the finite difference LES. The spectral LES displays a negative curvature in the time history of k. This is typical, and is due to the fact that the simulation begins with artificial initial conditions that do not have the proper phase relationships among the Fourier modes. It takes some time to establish the nonliner transfer of energy to the small scales, and this is reflected in the relatively slow decay at early times. The finite difference simulations on the other hand display a very rapid decay of TKE at early times. It is believed that this is due to the solution filtering. The initial spectrum (see the spectra results presented in the next section) is full, because this is a high Reynolds number case; however, the solution filtering instantly removes energy from the highest wavenumber range at the end of the first time step, and more energy is removed with each subsequent time step. A test without filtering is needed to be certain of what is happening, but the observed behavior seems consistent with this explanation.

The history of the density variance for the CBC case is given in Figure 11. The variance starts at zero since the initial conditions are incompressible. As with the TKE there is little difference among the simulations at later times, while the finite difference LES have lower fluctuation levels at early times compared to the spectral LES results.

The evolution of the three model coefficients for the CBC runs are shown in Figures 12-14. The values of C from both of the dynamic SGS model simulations shown in Figure 12 are larger than the constant value of 0.00846 used in the Smagorinsky model. The value for the finite difference LES is somewhat higher than that of the spectral LES.

The values of C_I are shown in Figure 13. In contrast to the Case 6 runs, the finite difference code computed a positive value of C_I. The values produced by the finite difference code are close to that assumed in the Smagorinsky model, while the spectral code has a much larger value of C_I. As pointed out in [12] the values of C_I are sensitive to the test filter, and for some filters a numerical instability can result. Since the test filter used in the finite difference code (see [10]) is not the same as the sharp cut-off filter used in the spectral calculations, it is not surprising that the values of C_I are different.

The turbulent Prandtl number, Pr_t, is shown in Figure 14. As with the Case 6 runs the dynamic SGS model gives values of Pr_t that are lower than the value used in the Smagorinsky model. The finite difference code produces somewhat lower values of Pr_t than the spectral code.

2.3   Simulation Results: Spectra

In order to get a better understanding of the behavior of the turbulence in the LES, three-dimensional velocity spectra, $E(\kappa)$, were computed. The spectrum represents how the energy in the velocity fluctuations is distributed over wavenumbers (length scales). The three-dimensional spectrum gives the energy within spherical shells in wave space, and κ is the magnitude of the wavenumber vector, $\vec{\kappa} = (\kappa_x, \kappa_y, \kappa_z)$. Figure 15 shows the spectra for the DNS run I2 and the two 32^3 finite difference LES runs I4 and I5 at time $t/\tau_0 = 0.2985$. There is very little difference between the two LES results, which indicates that there is little difference between the effect of the two SGS models for this case. In order to check whether the SGS is having any effect, a coarse grid DNS should also be run (on a 32^3 grid with no SGS model); however, this was not done. The LES spectra are close to that of the DNS for low wavenumbers; however, at higher wavenumbers there is significantly less energy in the LES spectra. This reduction in energy in the smaller scales results in the TKE for the LES being smaller than that in the DNS, as shown in Figure 3. It is believed that the reason the finite difference results are reduced for the higher wavenumbers is the solution filter that is applied in the finite difference code.

The spectra for the 48^3 LES are shown in Figure 16. The trends are similar to those seen for the 32^3 simulations; however, the reduction in the energy of the small scales is much less because of the greater resolution. This trend is consistent with the TKE seen in Figure 5.

The spectra for the CBC case are shown in Figures 17 and 18 at times $t/\tau_0 = 0.2870$ and 0.6612 respectively. As for the Case 6 LES, there is little difference between the results using the two SGS models. However, for this high Reynolds number case the difference between the finite difference LES spectra and the spectral LES code spectra are very large. The finite difference LES spectra are greatly distorted. The energy in the high wavenumbers (small scales) is reduced, while there is too much energy in the low wavenumbers (large scales). Even though the values of the TKE shown in Figure 10 are not that different among the simulations at the times when the spectra are computed, there is a significant difference in how the energy is distributed. This is a good indication that one cannot judge an LES solely by integrated quantities such as the TKE. It is important to consider spectra also.

It is believed that the behavior of the finite difference LES is due to the solution filtering, although further tests are needed to be certain of it effects. The filtering removes energy from the high wavenumber end of the spectrum. For this high Reynolds number case, where the spectrum is full (see the spectrum from the spectral code in Figures 17 and 18), the loss of energy in the higher wavenumber resolved scales affects the nonlinear transfer of energy from the lower wavenumber resolved scales. The nonlinear transfer of energy across some wavenumber, κ_c, occurs mostly between wavenumbers that lie close to κ_c. If the modes with wavenumber greater than κ_c are reduced in amplitude, then the nonlinear transfer of energy from wavenumbers smaller than κ_c are inhibited, and energy will tend to pile up in the wavenumbers just below κ_c. It is believed that this pile-up of energy is responsible for the excessive energy present in the low wavenumber region of the finite difference LES spectra. This possibility is explored further in the next section.

2.4 Discussion on the Effects of Filtering

Much of the discrepancy between the finite difference LES results and the spectral code LES results presented in the previous section seems to be due to the solution filtering procedure used in the finite difference code. In order to be certain of the effect, it is necessary to do simulations without the filter, or with different filter parameters. However, since such simulations have not been done at this time, our discussion must be restricted to the expected effect of the filter.

It should be clarified at this point that the decaying isotropic LES may be able to be done without solution filtering, or with filtering only applied infrequently. However, since filtering is needed in more complex flows for numerical stability, its effects should be fully understood. Examining the effects of filtering on isotropic turbulence may help with that understanding. Also, if the solution to the problem of the effect of filtering is to apply it only after a certain number of time steps, then the frequency of applying the filter becomes an ad hoc parameter, unless the frequency of application can somehow be related to the properties of the solution in a manner similar to how the model coefficients are determined in the dynamic SGS model.

The filter used for all of the current simulations has the filter parameter, α_f, set to $\alpha_f = 0.49$. The filter transfer function, $\widehat{G}(\kappa)$, is shown in Figure 19. With this value of the filter parameter, the filter is highly optimized. Therefore, the filter transfer function is fairly sharp, and its effect is mostly limited to a narrow high wavenumber region. Because the filter is applied at the end of each time step, it can have a cumulative effect. The other curves in Figure 19 show the effect of applying the filter multiple times. This is found simply by computing powers of the filter transfer function, $[\widehat{G}(\kappa)]^N$, where N is the number of times the

filter is applied. As one would expect, the more the filter is applied the greater its effect on the large scales, and the more the effective cutoff wavenumber moves to smaller wavenumbers.

For a linear problem with constant coefficients, the effect of the application of a solution filter can be predicted in the above way, and N is just the number of time steps. However, for nonlinear problems (such as turbulence simulations) or problems with nonconstant coefficients, energy is transferred among wavenumbers, so that in addition to the simple multiplicative effect of repeated application of the filter, the solution also depends on the transfer of energy from other wavenumbers. For decaying isotropic turbulence, applying the filter multiple times will greatly diminish the energy in the high wavenumber part of the spectrum, but the nonlinear transfer of energy will tend to feed energy back in. So the effective filter transfer function will be different than $[\widehat{G}(\kappa)]^N$, and in fact it may depend on wavenumber.

In order to make a rough estimate of the effective transfer function for the solution filtering procedure, the ratio of the spectrum from the Case 6 finite difference LES runs (using the dynamic SGS model) to the spectrum from the DNS run I2 is computed and shown in Figure 20 for the LES runs using a 32^3 grid and a 48^3 grid. It has a general shape similar to the filter transfer function of the solution filter applied many times. This is only a qualitative comparison, but it gives an indication that explanation for the depleted energy in the smaller resolved scales may be correct.

In Spyropoulos[11] the use of solution filtering with a compact finite difference scheme was examined (see Figure 3.20 therein). It was found that without filtering a 32^3 LES using a compact finite difference scheme had a pile-up of energy in the highest wavenumbers. LES were performed on a 48^3 grid with solution filtering using a sharp cut-off (SCO) filter, where the cut-off wavenumber was chosen to be 16 to make the simulation equivalent to a 32^3 simulation. The results showed excellent agreement with a 32^3 spectral LES. However, an important difference between the current filter used in FDL3DI and the SCO filter used in [11] is that the filter transfer function for the SCO filter is the same no matter how many times it is applied. For complex flow problems one cannot use the SCO filter, which is defined in Fourier wave space; one can only approximate it, as with the filter used in the current LES.

3 Commutation of Filtering and Differentiation

A fundamental question regarding LES was raised by Ghosal & Moin[6] who pointed out that for nonuniform grids the filtering operation does not commute with differentiation as is generally assumed when the LES equations are derived. They were able to devise a filtering method that commutes with differentiation up to second order in the filter width. However, this is a difficulty if one wishes to use high-order finite difference schemes, such as the compact differencing schemes used here. A filter which commutes to arbitrary order was developed by van der Ven[13]; however, it is limited to infinite domains with no boundaries. A survey of the issue and an attempt to create discrete filters which commute with discrete derivatives is given by Blaisdell[1].

In a recent paper Vasilyev[14] developed a set of filters which commute with finite difference schemes with an error of a certain order. The filter can be chosen so that the order of the commutation error is higher than the order of the truncation error of the difference scheme. In this way the commutation error will be negligible relative to the truncation error. The filters of Vasilyev are quite flexible and compact filters are available which have filter transfer function similar to those used in the current study.

As part of the current project the explicit (i.e. noncompact) filters of Vasilyev were tried on some test problems. Following [14] the test function is a 16th order Chebyshev polynomial defined on a stretched

hyperbolic tangent grid. The coefficients of the filters are given in Table I of the reference, although the table is incomplete and it was necessary to derive some of the needed coefficients. The finite difference scheme chosen is 4th-order central with 4th-order boundary closures. Figure 21 shows the results for the discretization (truncation) error (DE) and the commutation error (CE). The r.m.s. error is plotted, and the maximum error behaves similarly. The number after the CE indicates the order to which the filter is supposed to commute. The 2nd-order commuting filter gives an error (CE2) that is greater than the truncation error of the finite difference scheme. The 4th-order filter produces an error that is comparable to the truncation error. The 6th- and 8th-order commuting filters both give commutation errors that are significantly less than the truncation error.

Figure 21 can be compared to Figure 10 of the reference. Note that in Figure 21, for a large number of grid points, the commutation error actually scales with the order of the difference scheme rather than the advertised order. However, the level of the commutation error is significantly less than the truncation error, so that the difference in the scaling is inconsequential. In discussing the discrepancy with Dr. Vasilyev (private communication, 1998), it was found that the difference may be that he did not consider the commutation error coming from regions near the boundary in the test cases he ran, whereas the commutation error is greatest near boundaries due to the boundary treatments necessary in both the difference scheme and the filter.

In principle, the commuting filters of Vasilyev can be used so that commutation errors are not a concern. However, within the LES community this is not being done. The issue of commutation errors is basically being ignored. (Some researchers state that the commutation error is wrapped into the SGS model, but this amounts to ignoring the issue.) It may be that commutation errors are not significant. A few researchers that I have contacted informally have computed the commutation error from their LES data and found that it is small relative to the other terms in the LES equations. This is not quite the same as computing the commutation error based on fully resolved data, but it does give an indication of whether commutation errors should be of concern. In general, it is recommended that LES include computation of the commutation error as a check.

One point that needs to be kept in mind is that the manner in which filtering is used to derive the LES equations is not the same as the the solution filtering procedure applied in the current finite difference code. The difference between the two remains unclear at this point.

4   Conclusions and Suggestions for Future Work

The LES code, FDL3DI.LES, developed at the Air Force Research Lab has been applied to the problem of decaying compressible isotropic turbulence. The code uses a high-order compact finite difference scheme coupled with a high-order compact spatial filter. Comparisons were made with previous results from a Fourier spectral code.

The finite difference code was first tested on a DNS problem and excellent agreement was found with the results of the spectral code. LES were run for a low Reynolds number case, corresponding to the DNS test case, and a high Reynolds number case, corresponding to the experiment of Compte-Bellot and Corrsin. For the low Reynolds number case, the finite difference simulations on a 48^3 grid seem to agree well with the 32^3 LES from the spectral code. The turbulent kinetic energy and the density variance histories agree well. The values of the SGS model coefficients C and Pr_t are also similar between the two codes. The coefficient C_I differs between the two codes; however, it is noted that this coefficient is sensitive to the type of test

filter used to compute it, which may account for the difference. The 32^3 LES of this case showed too rapid a decay of the turbulent kinetic energy. Examination of spectra show that this behavior is due to a loss of energy from the smaller resolved scales. It is believed that this is due to the effect of the solution filter applied in the finite difference code, although it may also be that the 32^3 grid is too small to use with the finite difference method for this problem.

For the high Reynolds number case only 32^3 LES were performed. The history of the turbulent kinetic energy from the finite difference LES shows very rapid decay decay at early times, but shows good agreement with the spectral code LES results at later times. However, examination of the spectra reveals that the finite difference LES spectra are highly distorted, with too much energy in the large scales and too little energy in the small scales. Again, it is believed that this behavior is due to the solution filtering procedure, although further tests are needed to be certain. It is also likely that the finite difference code will require a finer grid in order to perform an adequate LES of this high Reynolds number case.

As part of the project, the issue of the commutation of filtering and differentiation was considered. The filters recently proposed by Vasilyev were considered. It was verified on a test problem that the higher order commuting filters produce a commutation error that is less than the truncation error of the finite difference scheme used. Although, these filters hold promise, they are not widely used. Also, there is an unsettled issue as to how the solution filtering procedure used here is related to the filtering assumed in the derivation of the LES equations, which is where the commutation issue arises.

There are several tasks that should be done to complete and extend the current work. First, the simulation cases considered did not include coarse grid DNS runs. These should be done in order to determine whether the SGS model has a significant effect or not. It also would give an indication of the effect of the numerical method. The simulations should be rerun without solution filtering, so that it can be determined whether the observed discrepancies between the finite difference LES and the spectral code LES are due to excessive filtering of the solution. Also, the Compte-Bellot and Corrsin case should be redone using 48^3 and a 64^3 grids.

An extension of the current research that should be useful is to investigate the effect of solution filtering in more detail. Different filter parameters or different types of filters can be used. The filter can be applied at various intervals instead of after each time step. The filter interval then becomes an additional parameter; however, it may be possible to find a rational way to determine the filtering frequency. Lastly, the relationship between solution filtering and the filtering assumed in deriving the LES equations should be considered.

References

[1] Blaisdell, G. A., "Commutation of Discrete Filters and Differential Operators for Large-Eddy Simulation," Advances in DNS/LES, Proceedings of the First AFOSR International Conference on Direct Numerical Simulation and Large Eddy Simulation, Louisiana Tech Univ., Ruston, LA, USA, August 4-8, 1997, pp. 333-340, Greyden Press, Columbus.

[2] Blaisdell, G. A., "A Review of Benchmark Flows for Large Eddy Simulation," Final Report for the AFOSR Summer Faculty Research Program, Wright Laboratory, August 1997.

[3] Compte-Bellot, G. and Corrsin, S., "Simple Eulerian Time Correlation of Full and Narrow Band Velocity Signals in Grid Generated, Isotropic Turbulence," *J. Fluid Mech.*, Vol. 48, pp. 273-337, 1971.

[4] Gaitonde, D. V., Shang, J. S., and Young, J. L., "Practical Aspects of High-Order Accurate Finite-Volume Schemes for Electromagnetics," AIAA Paper 97-0363, Jan. 1997.

[5] Gaitonde, D. V. and Visbal, M. R., "High-Order Schemes for Navier-Stokes Equations: Algorithm and Implementation into FDL3DI," AFRL-VA-WP-TR-1998-3060, Air Force Research Laboratory, Wright-Patterson AFB, OH, Aug. 1998.

[6] Ghosal, S. and Moin, P., "The basic equations for the large eddy simulation of turbulent flow in complex geometry," *J. Comp. Phys.*, Vol. 118, pp. 24–37, 1995.

[7] Lele, S. K., "Compact finite difference schemes with spectral-like resolution," *J. Comp. Phys.*, Vol. 103, pp. 16-42, 1992.

[8] Moin, P., Squires, K., Cabot, W., and Lee, S., "A Dynamic Subgrid-Scale Model for Compressible Turbulence and Scalar Transport," *Phys. Fluids A*, Vol. 3, pp. 2746-2757, November 1991.

[9] Ristorcelli, J. R. and Blaisdell, G. A., "Consistent Initial Conditions for the DNS of Compressible Turbulence," *Phys. Fluids*, Vol. 9, pp. 4-6, January 1997.

[10] Rizzetta, D. P., Visbal, M. R., and Blaisdell, G. A., "Application of a High-Order Compact Difference Scheme to Large-Eddy and Direct Numerical Simulation," AIAA Paper 99-3714, June 1999.

[11] Spyropoulos, E. T., "On Dynamic Subgrid-Scale Modeling for Large-Eddy Simulation of Compressible Turbulent Flows," Ph.D. Thesis, School of Aeronautics and Astronautics, Purdue Univ., West Lafayette, Indiana, December 1996.

[12] Spyropoulos, E. T. and Blaisdell, G. A., "Evaluation of the Dynamic Model for Simulations of Compressible Decaying Isotropic Turbulence," *AIAA J.*, Vol. 34, pp. 990-998, May 1996.

[13] van der Ven, H., "A family of large eddy simulation (LES) filters with non-uniform filter widths," *Phys. Fluids*, Vol. 7, pp. 1171–1172, 1995.

[14] Vasilyev, Oleg V., Lund, Thomas S., Moin, Parviz, "A General Class of Commutative Filters for LES in Complex Geometries," *J. Comp. Phys.*, Vol. 146, No. 1, pp. 82-104, 1998.

[15] Visbal, M. R. and Gaitonde, D. V., "High Order Accurate Methods for Unsteady Vortical Flows on Curvilinear Meshes," AIAA Paper 98-0131, Jan. 1998.

Table 1: Case descriptions for decaying isotropic turbulence simulations.

Run	Case	Code	DNS/LES	SGS Model	Grid
I1	Case 6	spectral	DNS	none	128^3
I2	Case 6	FDL3DI	DNS	none	128^3
I3	Case 6	spectral	LES	dynamic	32^3
I4	Case 6	FDL3DI	LES	dynamic	32^3
I5	Case 6	FDL3DI	LES	Smagorinsky	32^3
I6	Case 6	FDL3DI	LES	dynamic	48^3
I7	Case 6	FDL3DI	LES	Smagorinsky	48^3
I8	CBC	spectral	LES	dynamic	32^3
I9	CBC	FDL3DI	LES	dynamic	32^3
I10	CBC	FDL3DI	LES	Smagorinsky	32^3
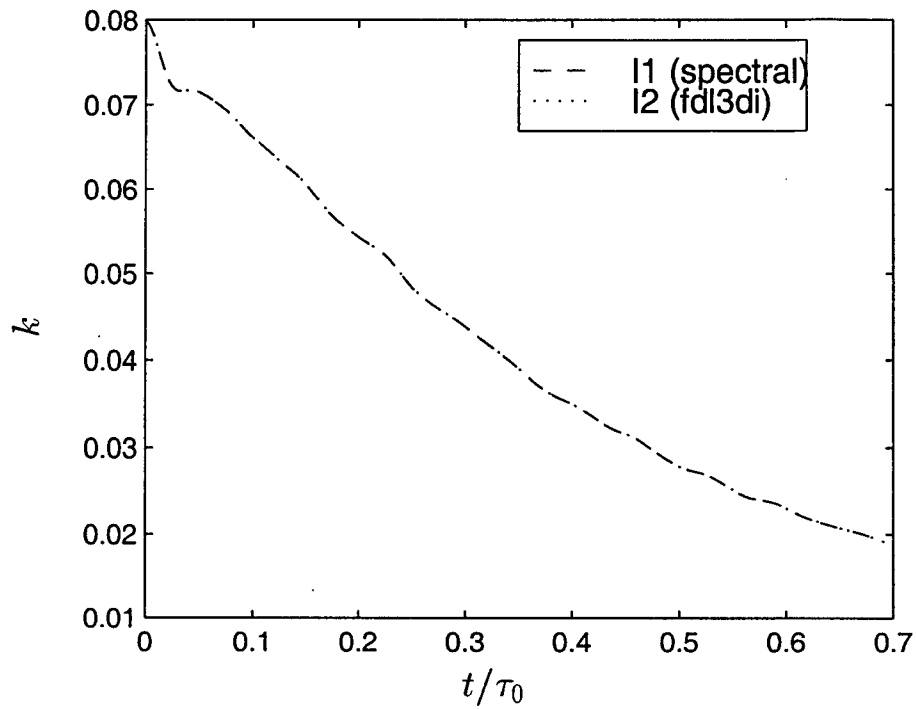
Figure 1: Evolution of the turbulent kinetic energy for the DNS runs, I1 and I2.
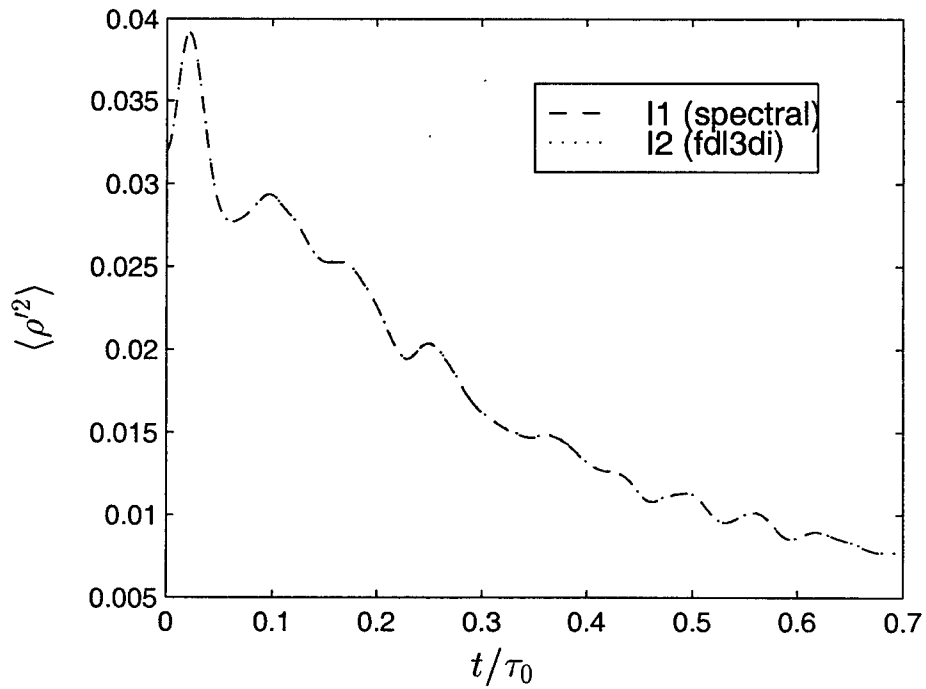
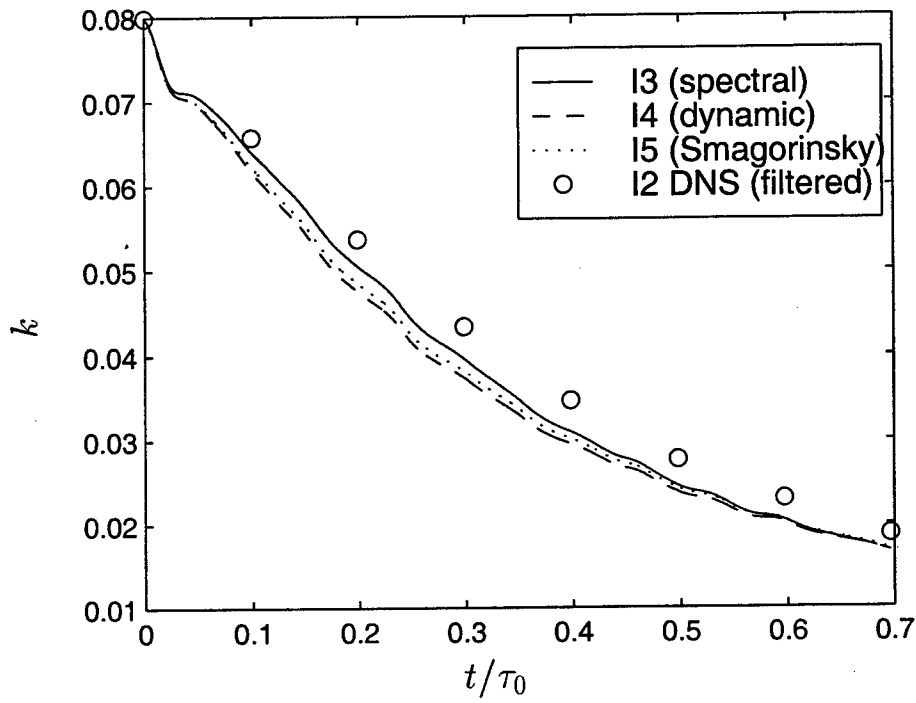Figure 2: Evolution of the density variance for the DNS runs, I1 and I2.

Figure 3: Evolution of the turbulent kinetic energy for the 32^3 LES runs, I3, I4 and I5 compared to the filtered DNS data from run I2.
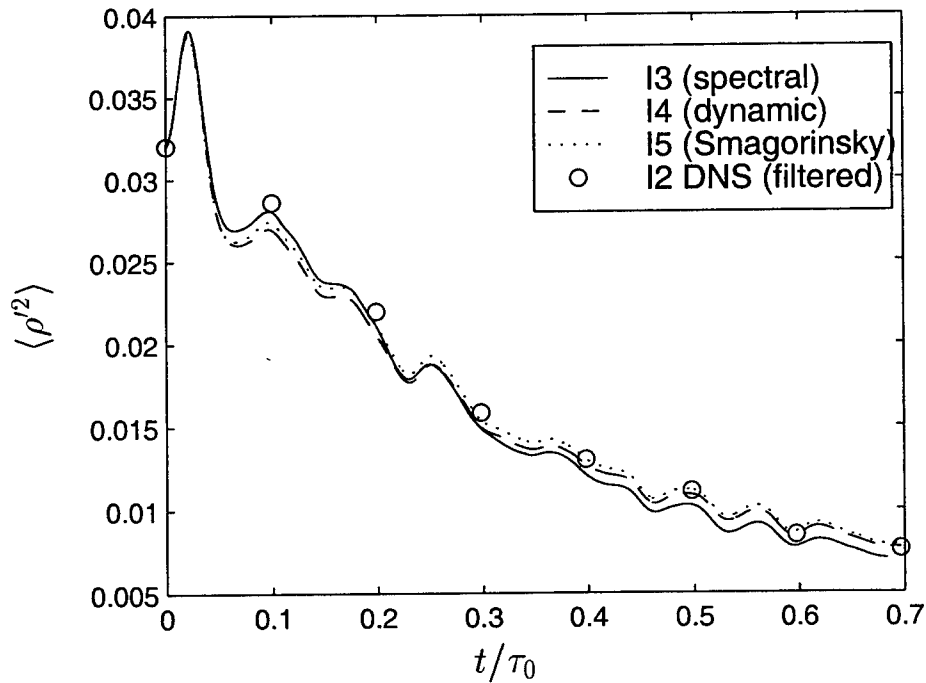

Figure 4: Evolution of the density variance for the 32^3 LES runs, I3, I4 and I5 compared to the filtered DNS data from run I2.
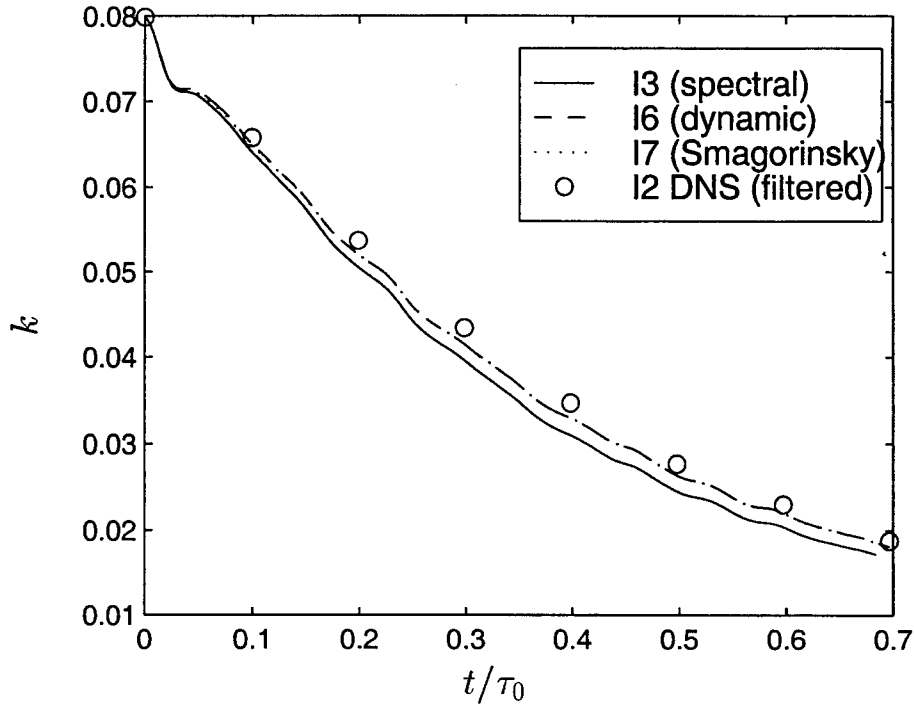
Figure 5: Evolution of the turbulent kinetic energy for the 48^3 LES runs, I6 and I7, and the 32^3 LES run I3 compared to the filtered DNS data from run I2.
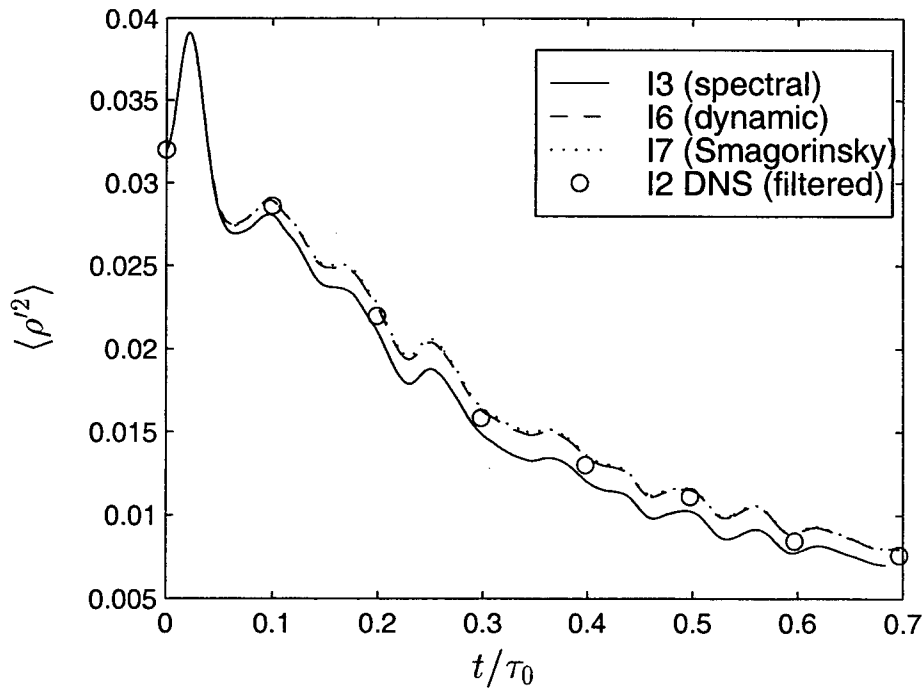

Figure 6: Evolution of the density variance for the 48^3 LES runs, I6 and I7, and the 32^3 LES run I3 compared to the filtered DNS data from run I2.
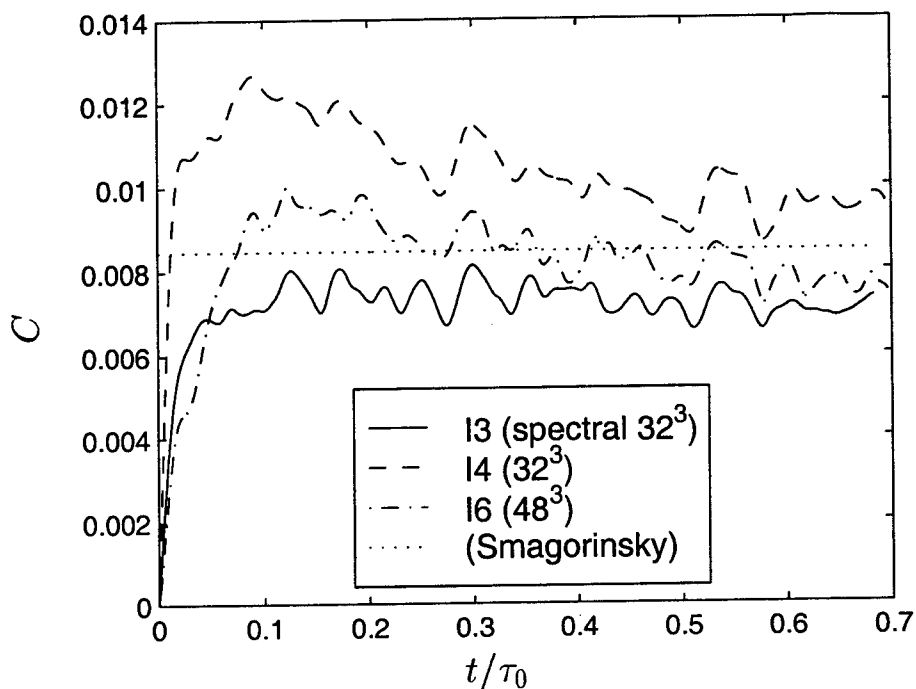
Figure 7: Evolution of the model coefficient C for the dynamic Case 6 LES runs, I3, I4 and I6, compared to the constant value used in the Smagorinsky model.
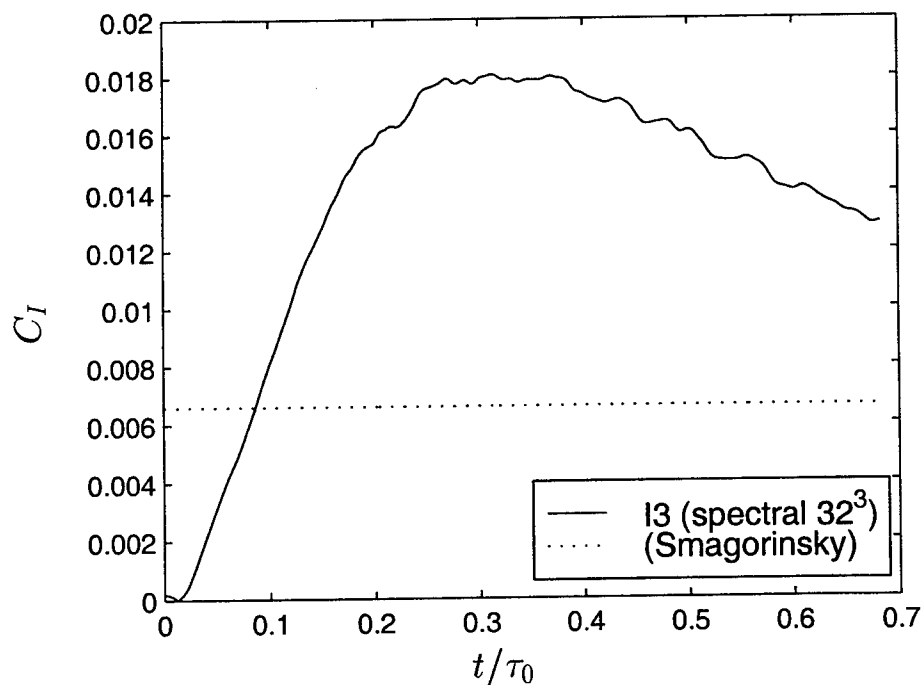

Figure 8: Evolution of the model coefficient C_I for the dynamic Case 6 LES run, I3, compared to the constant value used in the Smagorinsky model. (The values for runs I4 and I6 are zero.)
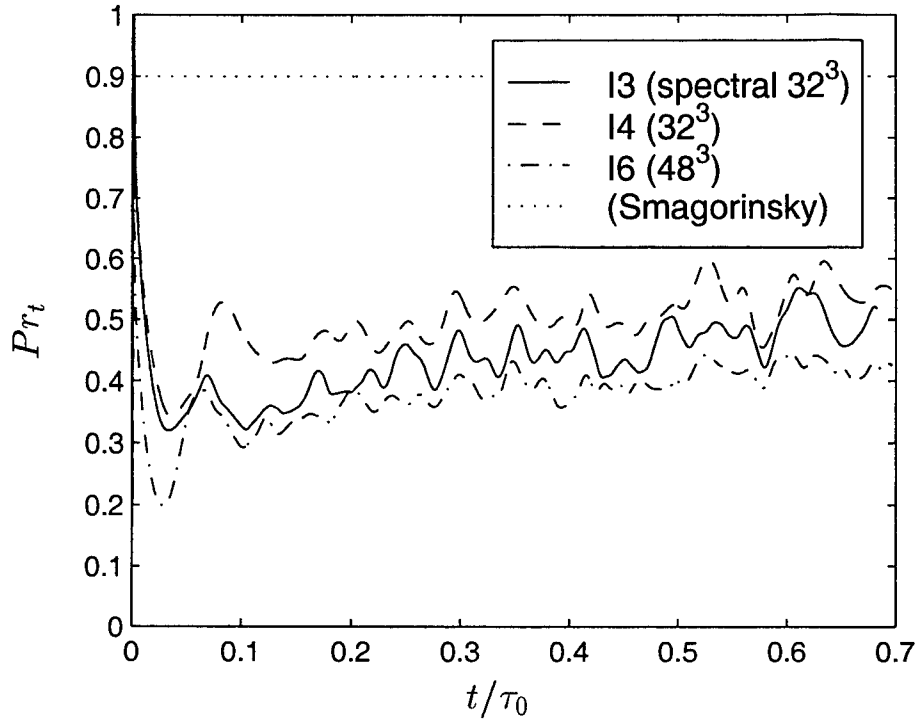
Figure 9: Evolution of the turbulent Prandtl number Pr_t for the dynamic Case 6 LES runs, I3, I4 and I6, compared to the constant value used in the Smagorinsky model.
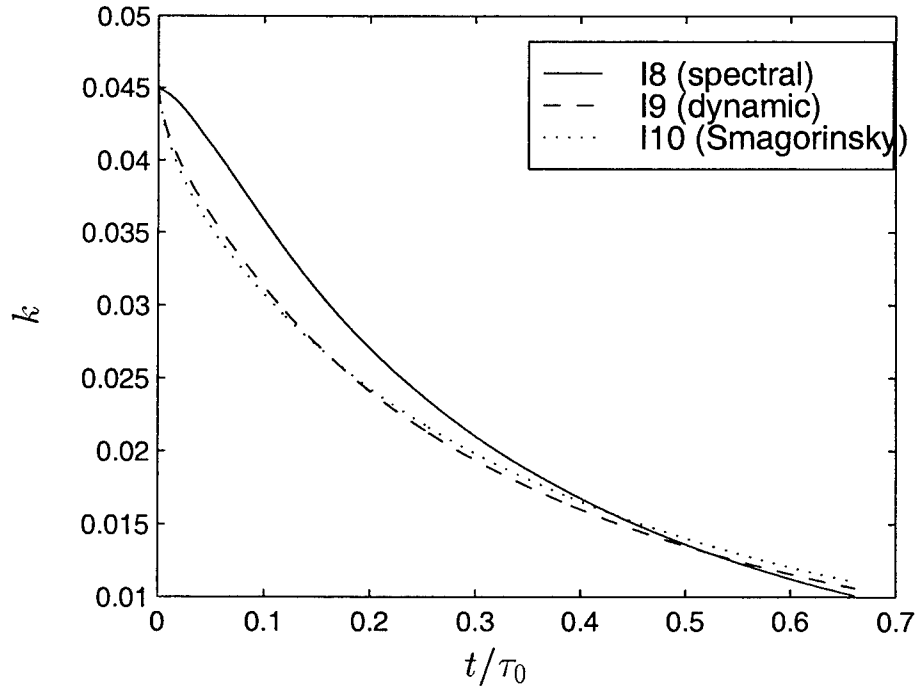

Figure 10: Evolution of the turbulent kinetic energy for the CBC LES runs, I8, I9 and I10.
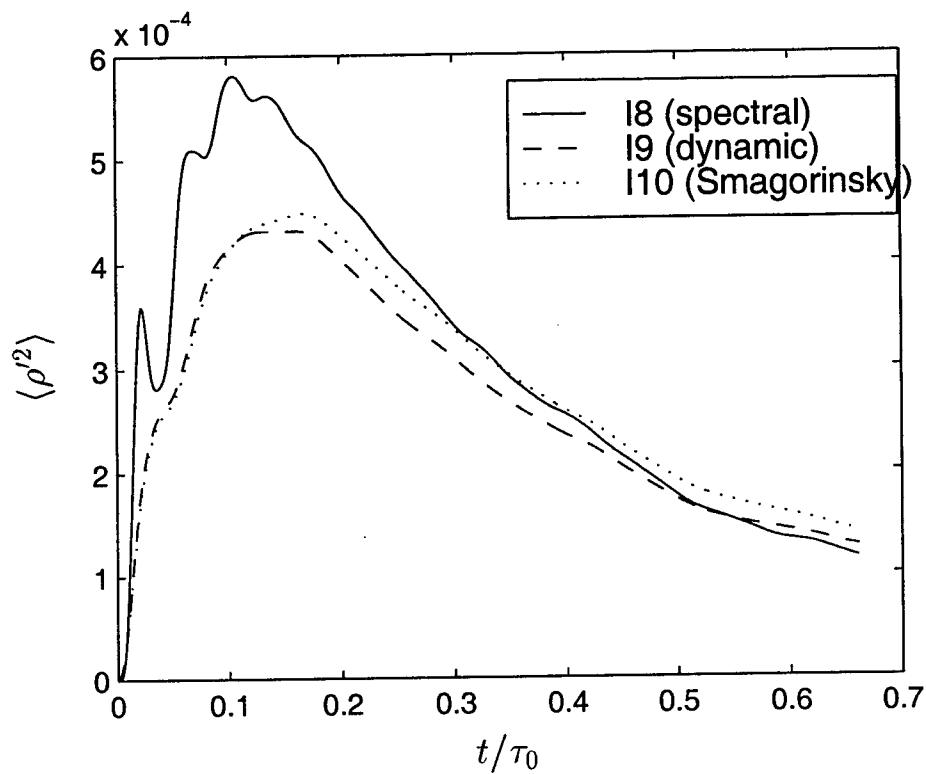
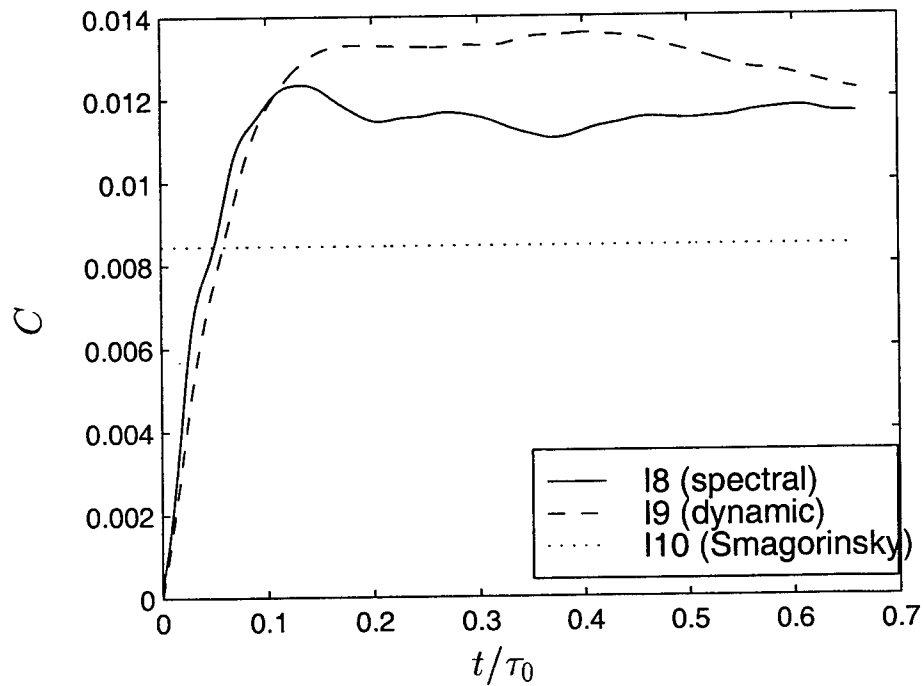Figure 11: Evolution of the density variance for the CBC LES runs, I8, I9 and I10.

Figure 12: Evolution of the model coefficient C for the dynamic CBC LES runs, I8, I9 and I10, compared to the constant value used in the Smagorinsky model.
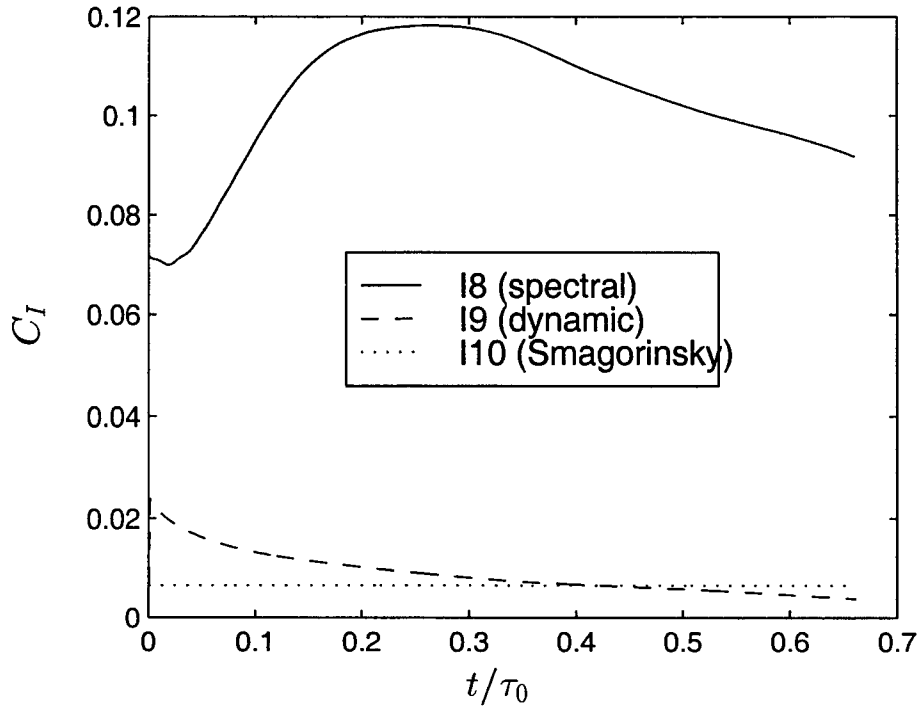
Figure 13: Evolution of the model coefficient C_I for the dynamic CBC LES runs, I8, I9 and I10, compared to the constant value used in the Smagorinsky model.
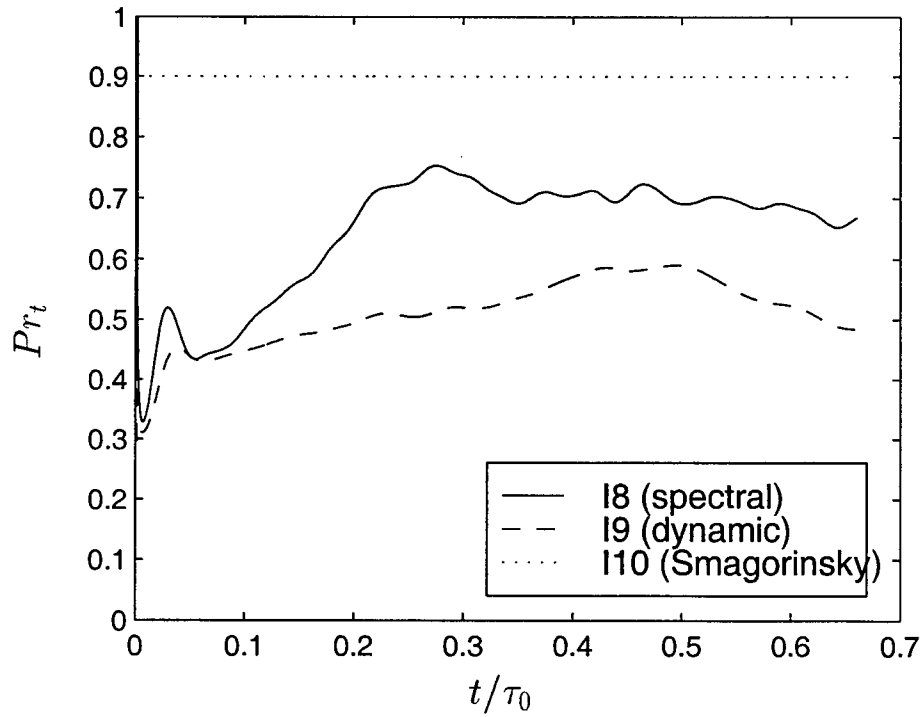
Figure 14: Evolution of the turbulent Prandtl number Pr_t for the dynamic CBC LES runs, I8, I9 and I10, compared to the constant value used in the Smagorinsky model.
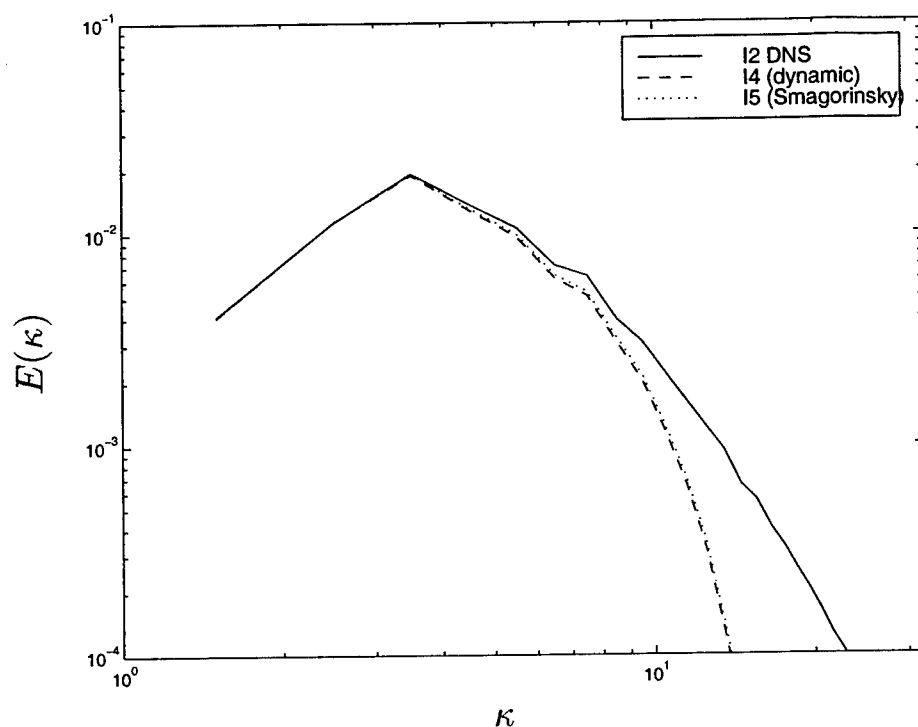
Figure 15: Three-dimensional velocity spectrum for the 32^3 Case 6 LES, I4 and I5, compared with the DNS I2 at time $t/\tau_0 = 0.2985$.
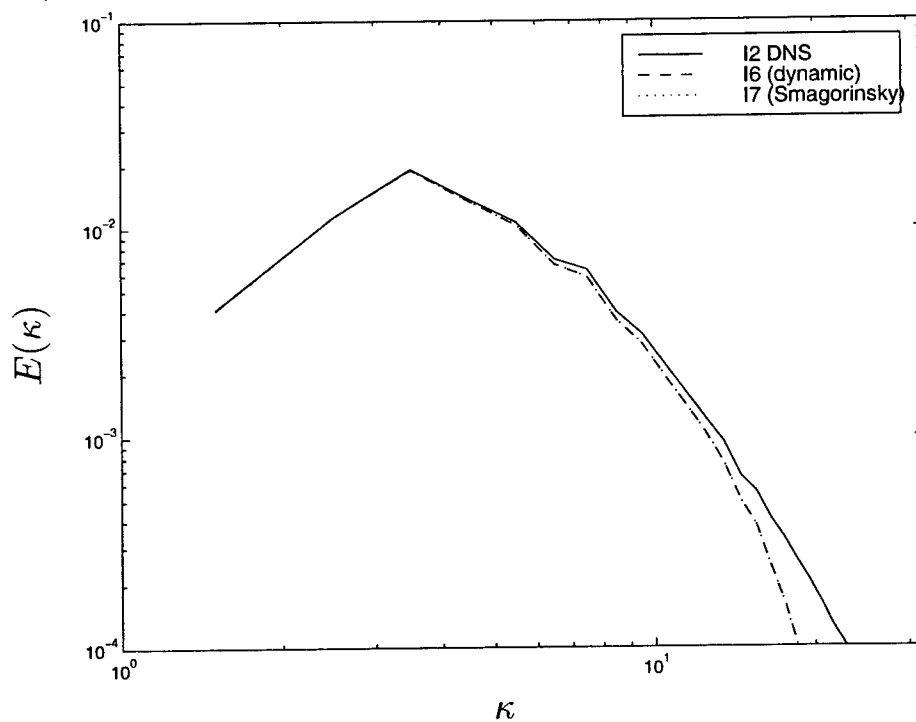
Figure 16: Three-dimensional velocity spectrum for the 48^3 Case 6 LES, I6 and I7, compared with the DNS I2 at time $t/\tau_0 = 0.2985$.
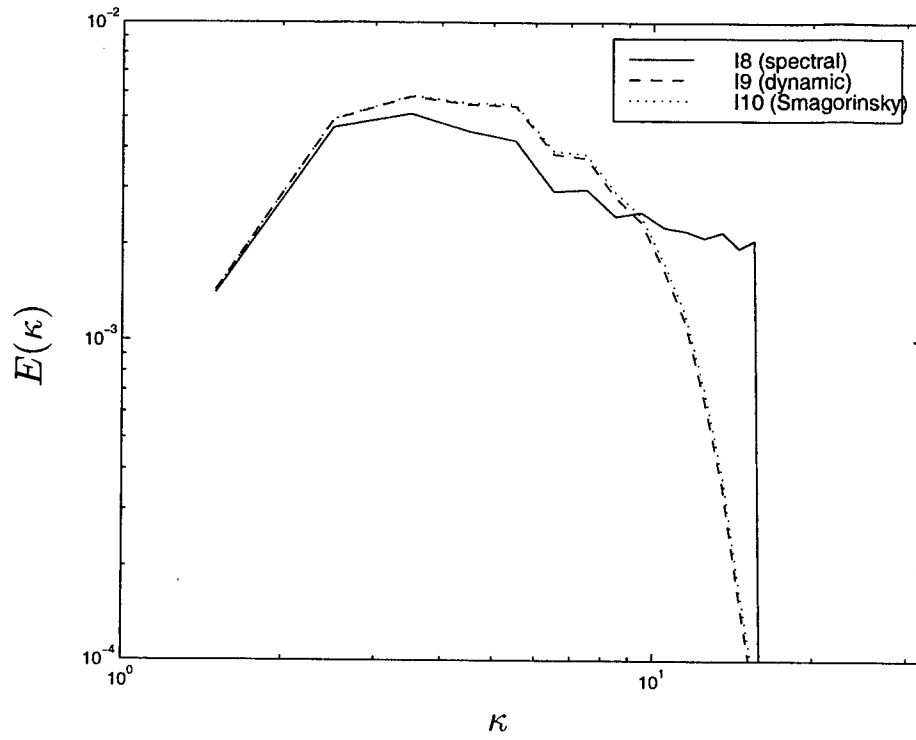
Figure 17: Three-dimensional velocity spectrum for the CBC LES, I9 and I10, compared with the spectral code LES I8 at time $t/\tau_0 = 0.2870$.
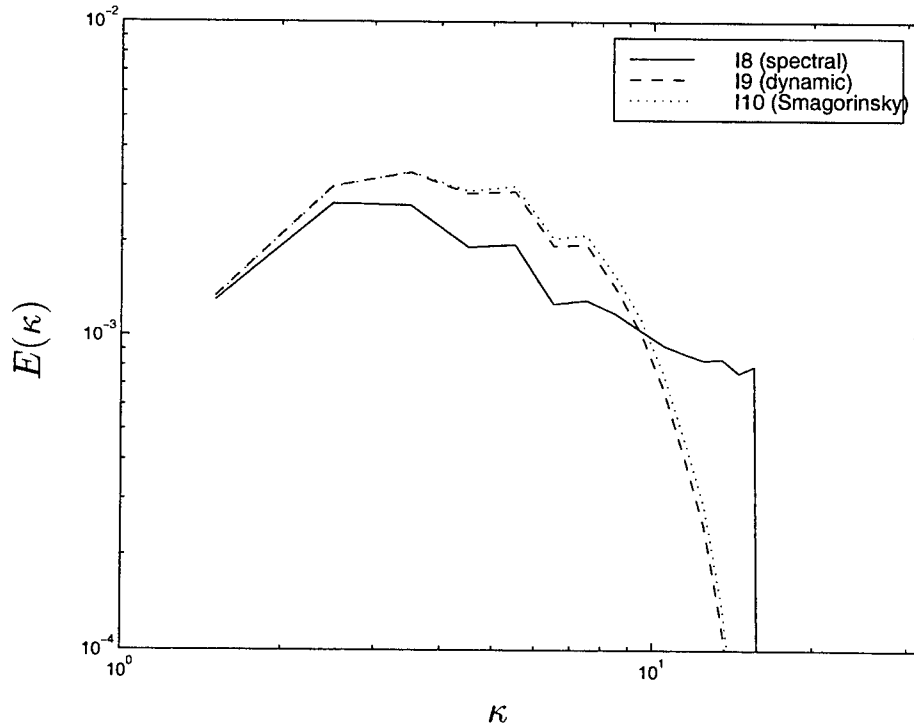
Figure 18: Three-dimensional velocity spectrum for the CBC LES, I9 and I10, compared with the spectral code LES I8 at time $t/\tau_0 = 0.6612$.
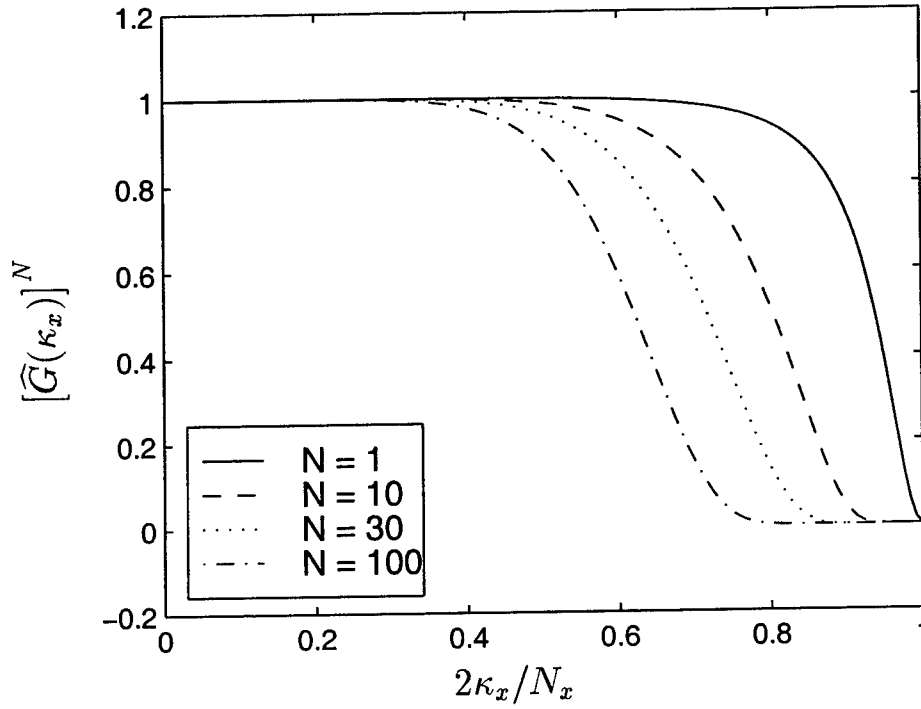
Figure 19: Filter transfer function for the solution filter used in FDL3DI (with $\alpha_f = 0.49$) applied multiple times.
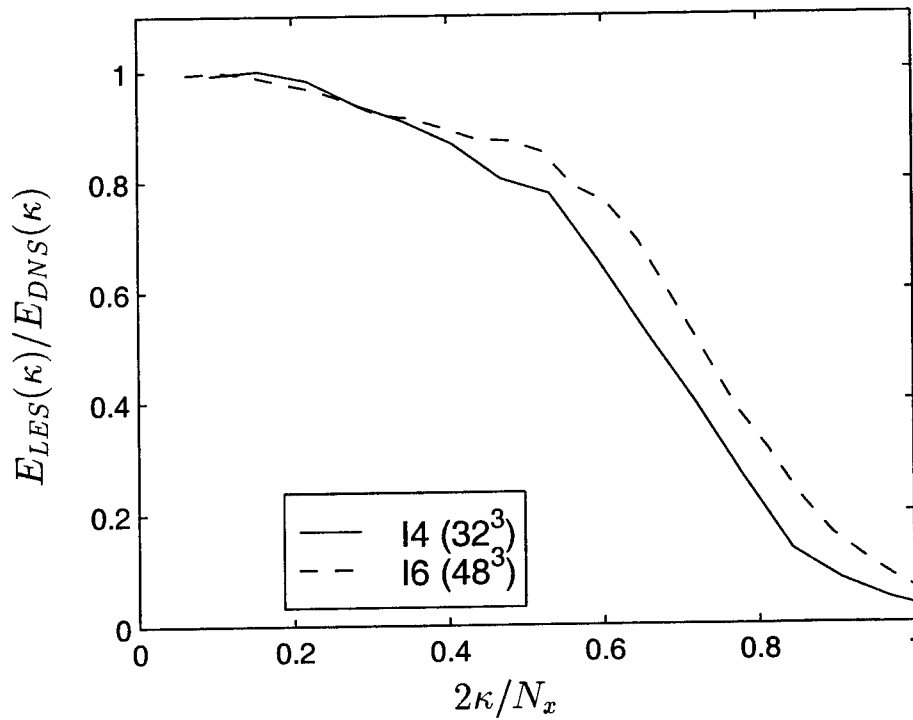
Figure 20: Ratio of the three-dimensional velocity spectrum from the Case 6 LES, I4 and I6, to that from the DNS I2.
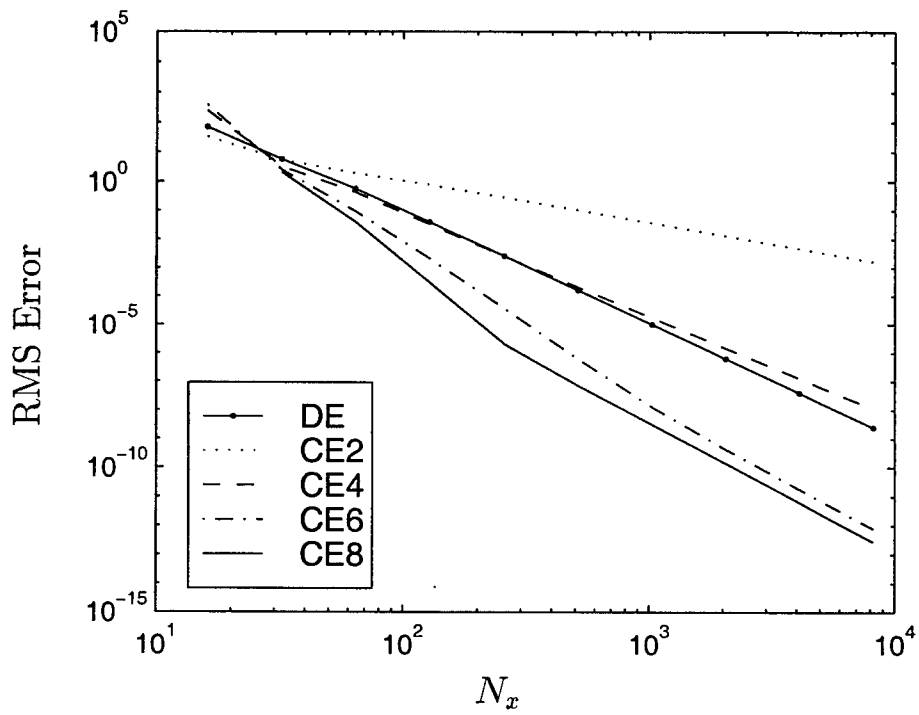
Figure 21: Comparison of the r.m.s. discretization error (DE) for a 4th-order finite difference scheme with the commutation error (CE) for commuting filters of various orders.

AN INSECT-INSPIRED MODEL FOR
OPTIC FLOW PROCESSING HARDWARE

John K. Douglass
Division of Neurobiology

University of Arizona
611 Gould-Simpson Bldg.
Tucson, AZ 85721

Abstract

Many animals use panoramic visual motion cues to monitor their progress through the world and correct deviations from their intended courses. A computational model inspired by features of insect visual systems was developed to examine the roles of both physiological tuning properties and network architecture in generating circuits that are selective for optic flow field type and position. Arrays of small-field motion detectors (SFMDs) endowed with various functional properties (selectivity to motion direction and speed; purely excitatory outputs or excitatory/inhibitory outputs) were tested with alternative spatial synaptic output patterns to evaluate what features of SFMD outputs are most important for generating wide-field optic flow selectivity. The biomimetic flow-selective properties exhibited by this model include "coarse coding" for optic flow type, maintenance of selectivity for optic flow type across a range of optic flow speeds, and sensitivity to optic flow position relative to receptive field center of a flow-selective collator neuron or summating element. The key design features are a simple feed-forward network composed of two basic stages: (a) arrays of sfmds that respond according to temporal delays determined by motion direction and speed, and (b) simple spatial patterns of purely excitatory synaptic connection strengths that can be designed to generate selectivity for a wide variety of flow fields. Flow field selectivity can be produced by optimizing either of two basic design features: the response properties of the sfmds and the spatial synaptic output mappings to the collator stage. Design parameters are flexible: if roughly optimal values for either of the two basic features are met, robust model behavior is obtained without the need for precise control of the other parameters. The results of this investigation suggest specific design strategies for man-made optic flow analyzers. These strategies are applicable to flow analysis in

general, and therefore can be extended to other types of spatial processing tasks, including tracking and pursuit of small mobile targets.

Introduction

This report describes a computational model for optic flow processing that was inspired by known features of motion processing circuitry in the visual systems of flies (dipteran insects). The model was developed as a result of the author's participation in the AFOSR summer faculty research program at Eglin AFB, FL, where several basic information processing strategies employed by insect visual systems were identified for possible applications to machine vision hardware (Douglass, 1997). In particular, the optic flow processing model described below was designed to elucidate a basic issue in neuronal information processing: to what extent are neuronal circuits based on broad parametric tuning (coarse-coding) mechanisms as opposed to sharp tuning, or "matched filtering"? Previous theoretical work based on insect circuits that process directional wind inputs has suggested that directional information transfer is maximized when the direction-sensitive circuit elements exhibit a single optimal tuning curve bandwidth (Theunissen and Miller, 1991). The model visual processing circuit presented here challenges this view, by demonstrating that the components of the circuit need not be restricted to a single directional tuning curve width in order to generate selectivity for optic flow type and position.

This model is designed to examine the properties of alternative configurations of idealized optic flow processing circuitry. Although it is based on known features of insect visual pathways, many relevant details of this system are unknown or only inferred at present. Thus, the goal is not to faithfully imitate the actual circuitry, but to generate testable predictions and

elucidate important general requirements for optic flow processing. The results of this investigation suggest specific design strategies for man-made optic flow analyzers. These strategies are applicable to flow analysis in general, and therefore can be extended to other types of spatial processing tasks, including tracking and pursuit of small mobile targets.

Biological Bases for the model. Conceptual models of biological optic flow processing mechanisms generally fall into two categories, based on the types of input features that are used to generate flow field selectivity. These two alternative categories are known as the "direction mosaic hypothesis" and the "vector field hypothesis" (Duffy and Wurtz, 1991). The direction mosaic hypothesis postulates an array or mosaic of small-receptive field neurons or circuits, each small-field unit being sensitive to local motion direction and having its outputs arranged to yield large-field selectivity for a particular type of optic flow. This type of mechanism should inherently be sensitive to the positions of circular or radial flow fields with respect to the receptive field of the circuit. In contrast, the vector field hypothesis relies on responsivenes to fundamental features of particular types of optic flow, features that are distributed throughout the flow field and often are less position-dependent. These features include the divergence among local vector directions in radial fields, the curvature, or "curl" of circular fields, and the gradient in local vector magnitudes that is present within unidirectional, or "planar" flow fields (c.f. Koenderink, 1986).

Although there is some evidence for the operation of both the direction mosaic and vector field strategies in primate cortical circuitry (Duffy and Wurtz, 1991), the evidence from insects favors the direction mosaic hypothesis (Krapp and Hengstenberg, 1997). In both insects and vertebrates, various lines of evidence suggest that the fundamental building blocks for optic

5-4

flow analysis comprise two basic stages. The first stage consists of retinotopically mapped arrays of small-field-of-regard, direction-sensitive motion detectors, which obtain their inputs from an photoreceptors and peripheral visual interneurons, and which provide their outputs, in the second stage, to wide-field-of-regard collator neurons (c.f. Douglass and Strausfeld 1995; Morrone et al., 1995). The model described here also contains two basic stages, and is based on the direction mosaic hypothesis.

Several additional features of the model are based on known features of insect visual processing circuitry. In flies, a retinotopically mapped array of small-field neurons known as T5 cells corresponds to the small-field motion detectors in the 1st stage of the model. Intracellular recordings from identified T5 neurons have demonstrated their directional selectivity to motion (Douglass and Strausfeld, 1995), and a number of investigations (reviewed by Strausfeld, 1989; Strausfeld and Lee, 1991) show that they provide major inputs to wide-field collator neurons in a visual processing center known as the lobula plate. Many of the wide-field collators are sensitive to directional motion, and at least a subset is selective for complex optic flow fields (Krapp and Hengstenberg, 1996, 1997). There is strong anatomical evidence for the existence of four superimposed arrays of T5 cells within each retinotopic column, or visual sampling point (Strausfeld and Lee, 1991), thus, the model below specifies four arrays of small-field motion detectors (sfmds). Each array of sfmds produces maximal responses during stimulation by motion in a single "preferred direction" (PD), and the four PDs are oriented at 90-degree intervals. In flies, individual T5 neurons tend to have their outputs segregated into one of four distinct anatomical layers in the lobula plate (Fischbach and Dittrich, 1989; Strausfeld, 1989) that are individually sensitive to progressive, regressive, upward and downward motion

(Buchner et al., 1984). Both the physiological responses and anatomical stratification of various identified neurons with lobula plate inputs are consistent with a four-layered functional organization of the lobula plate (e.g. Eckert 1982; Douglass and Strausfeld 1996, 1998).

Methods

The model defines a feed-forward network with optic flow fields as its inputs, and responses of wide-field collator neurons as its outputs. The source code was written in Turbo Pascal (Borland International, Inc.). Based on the user's selection of model parameters, the model first defines a particular optic flow field and flow field position. It then computes responses of arrays of local motion detectors (sfmds); these responses serve as the inputs to model wide-field collator neurons that are designed to exhibit selectivity for a particular type of non-uniform optic flow (expansion, contraction, clockwise (CLW) rotation, counterclockwise (CCLW) rotation) based on a spatial pattern of synaptic output connection strengths from the sfmds.

Stimuli. The optic flow stimulus (Fig. 1, at left) is defined as a spatial array of local motion vectors. Each vector is defined by an (x,y) location within a plane orthogonal to the line of sight of a viewer, an eye, or a visual sensor, and by a local motion amplitude r and direction θ. Flow fields can be specified as one of five cardinal flow field types (pure clockwise rotation, counerclockwise rotation, expansion, contraction, or unidirectional flow), or as combinations of these cardinal types. In addition, the location of a flow field can be varied in relation to the receptive field center and receptive field diameter of the collator neuron. In its current implementation, local flow vectors are represented as their projections from 3-dimensional space

5-6

Fig. 1
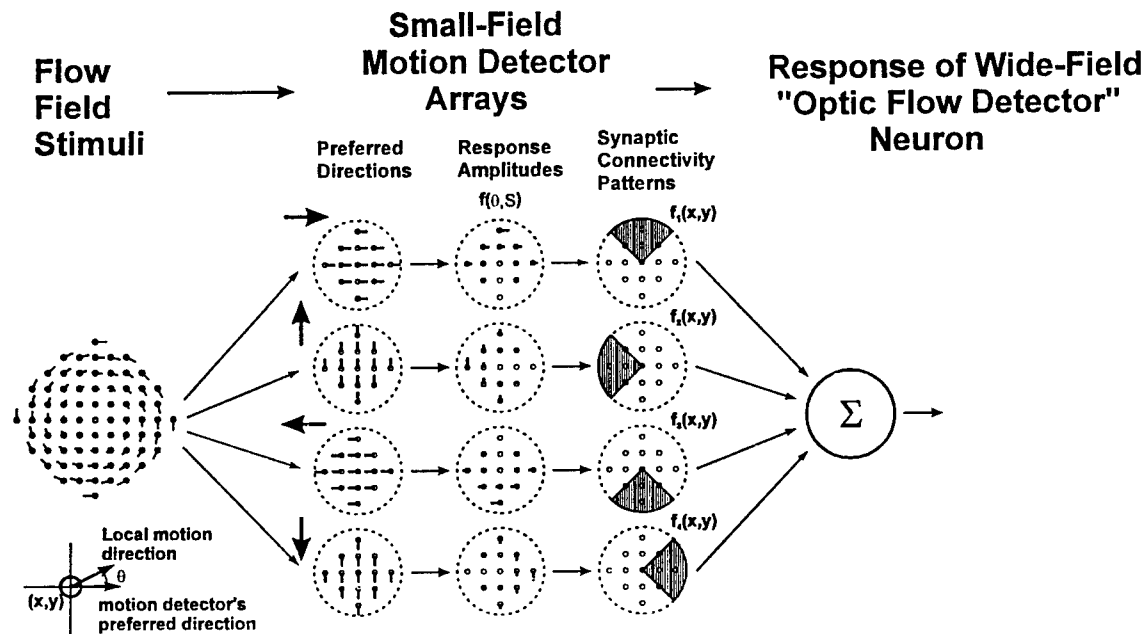
onto a 2-dimensional plane. This method is reasonably accurate for modelling moderately wide

visual fields that approach hemispherical views; in future implementations, the stimulus

definitions will be generalized to encompass the nearly panoramic fields of view that insects

possess and which are desirable to emulate in hardware.

For each sfmd within the receptive field of the collator neuron, the model defines a

stimulus vector direction (θ) and magnitude A according to (1) the type of optic flow field, and

(2) the distance r from the flow field center to the sfmd sampling point in visual space (This is an

approximation to the small region of visual space which is actually sampled by the small-field

inputs to an sfmd circuit; see below). The local stimulus direction θ, in degrees, is defined as the

difference between the local flow direction and the preferred direction (**PD**) of the sfmd:

$$\theta = \arcsin(y/r) - (PD+90) + c, \qquad xi > 0 \qquad\qquad (1)$$

$$\theta = (PD+90) + c - \arcsin(y/r), \qquad xi <= 0.$$

Xi and **y** are the horizontal and vertical components, respectively, of **r**, and the value of **c** determines the type of optic flow field (c = 0 for pure clockwise rotation, c = 90 for pure expansion (looming), and so on).

As illustrated at the left of Fig. 1, the local flow field vector magnitude **A** is proportional to its distance **R** from the optic flow field center and a flow field speed factor **k**. (K is translatable to rpm for rotational fields; deg./s for expanding or contracting fields):

$$A = kR \qquad\qquad (2)$$

Small-field motion detectors.

The small-field motion detectors (sfmds) depicted in Figure 1 were modeled in two ways. First, because appropriately detailed information on the response properties of actual sfmds were unavailable, a detailed circuit model of insect sfmds was used to generate responses to motion direction and speed that can be considered biologically plausible. Second, a purely analytical model was developed that captured the major features of the detailed model, yet could be conveniently altered in predictable ways. The analytical model was used as the first stage in the wide-field optic flow model of Figure 1.

Detailed SFMD Model based on Hassenstein-Reichardt Correlation -type Detectors.

The foundation for the more-detailed sfmd model is the classical Hassenstein-Reichardt correlation-type elementary motion detector (emd) circuit (Hassenstein and Reichardt, 1956;

Borst and Egelhaaf, 1989), which employs two light-sensitive input channels with slightly

different viewing angles, a fixed internal temporal delay or "circuit delay" in one channel, and a

mathematical multiplication of the outputs from the two channels. In the present implementation

(Fig. 2), a single edge motion stimulus is modeled as the sequential arrival of an increase or

decrease in light intensity at the two channels, the external delay in the stimulus arrival at the

two channels being determined by the motion speed and the angular distance between the

channels. The responses of each channel to changes in light intensity are defined using

exponential functions of time (J.K. Douglass, unpublished) that are designed to mimic typical

intracellular voltage responses to flicker observed in intracellular recordings from various

Fig. 2
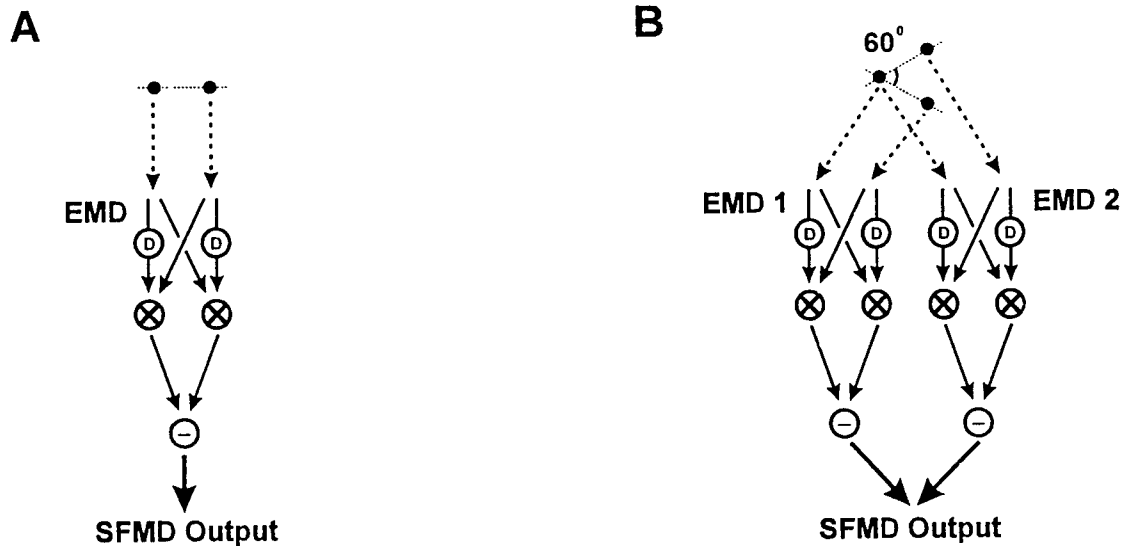
peripheral, non-motion-selective visual interneurons of calliphorid flies (e.g. Laughlin, 1981; Douglass and Strausfeld, 1995). Beyond the level of the photoreceptors, these responses typically consist of a hyperpolarization to "light ON" and a depolarization to "light OFF," and both responses are characterized by a rapid onset and relatively slow decay toward the baseline voltage.

The inputs to neighboring emd channels were defined to be 2^0 apart, and the fixed circuit delay was set at 12 ms, with an additional cable conduction delay of 8 ms within each channel. The emd response was defined as a non-spiking intracellular voltage, integrated over an interval of several ms after the arrival of a single edge at either channel. In order to produce both positive (excitatory) and negative (inhibitory) responses, the outputs from two mirror-symmetrical *emds* having opposite "preferred motion direction axes" were subtracted (Borst and Egelhaaf, 1990), and this result was taken to represent the response of a single *sfmd*. The preferred/null axis of this type of emd is the line defined by the two input channels. Motion of an edge that is orthogonal to the preferred/null axis produces a direction-dependent response (Grzywacz et al., 199). An alternative circuit configuration was also tested to illustrate the motion responses of a more spatially-distributed circuit (see Fig. 2B and Results), in which the outputs from two pairs of emds, their preferred/null axes separated by 60^0 (Buchner, 1976), were summed to produce the response of an sfmd.

Analytical SFMD Model.

The first stage of the optic flow model consists of sfmd responses to local motion vectors. The responses of each sfmd were defined as a function of motion direction (θ) and speed (S). Responses to motion speed were modelled as an exponential function,

$$r(S) = k*S* e^{(1-k*S)} .$$ (3)

The shape of this function (see Results, Fig. 5E) provides a good qualitative fit to speed response data that have been obtained during intracellular recordings from wide-field collator neurons in the lobula plate of flies (Eckert 1980, 1982; Hausen 1982; Hengstenberg 1982), as well as single-unit recordings from directional motion - sensitive neurons in area MT of the visual cortex of macaque monkeys (Maunsell and Van Essen, 1983). This approximate shape also is predicted from sfmd models during motion along the preferred/null axis (Results, Fig. 3).

To investigate the effects of sfmd directional tuning properties on the model output, the speed response function was convolved with one of four alternative cosine-shaped directional tuning functions $f(\theta)$ (Results, Fig. 5 A-D), chosen to encompass a broad range of possible characteristics of biological sfmds. The alternative directional functions are either broadly ($f_A(\theta)$, $f_B(\theta)$), or narrowly ($f_C(\theta)$, $f_D(\theta)$) tuned, and produce either monophasic (purely excitatory; $f_A(\theta)$, $f_C(\theta)$) or biphasic (both excitatory and inhibitory, $f_B(\theta)$, $f_D(\theta)$) responses as a function of motion direction. Identified sfmd output neurons in the fly brain (T5 cells, Douglass and Strausfeld, 1995) exhibit broad, combined excitatory/inhibitory responses to motion. Monophasic (e.g. purely excitatory) responses are also possible, as occurs in spiking neurons with little or no background activity, and as is also predicted for sfmd circuits that lack a subtraction stage, or in which the integration stage involves a biophysical mechanism known as shunting inhibition. The general equation used to specify the directional tuning of the sfmds is

$$R(\theta) = b + \cos [a*(\theta - (PD)], \quad |\theta - PD| < \pi/a,$$ (4)

$$R(\theta) = b, \quad \pi/a <= |\theta - PD| < \pi/a,$$

where R is the response as a function of motion direction, "b" specifies a constant offset from the zero response level, "a" determines the sharpness (or coarseness) of the directional tuning, and **PD** = the preferred direction of the sfmd, defined as the motion direction that produces the maximal response amplitude. The differences among the four alternative sfmd tuning functions in Figure 5 reflect the values that were chosen for parameters **a** and **b**:

function	a	b
f_A	1	0.5
f_B	1	0
f_C	2	0.5
f_D	2	0

In summary, the response of each sfmd was computed as a function of the speed and direction of the local motion vector at its location within a retinotopic map. The overall response of each sfmd is thus:

$$R(\theta,S) = R(\theta) * R(S) \tag{5}$$

Wide-field collator responses

Responses of a wide-field collator neuron were calculated by multiplying each sfmd output with a spatial synaptic weight function, then computing the linear sum of all synaptic outputs. Simple, "uniform" output mappings to a collator neuron (Fig. 3, upper portion) were based on those proposed by Tanaka et al. (1989), and were designed to result in model collator neurons that are selective for one of four cardinal flow field types: clockwise rotation, counterclockwise rotation, expansion (positive looming) and contraction (negative looming). Also tested were cosine-shaped "gradient" output mappings (Fig. 3, lower portion) similar to

Fig. 3

Uniform Maps
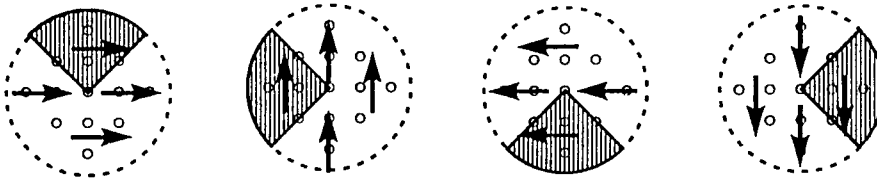
Gradient Maps
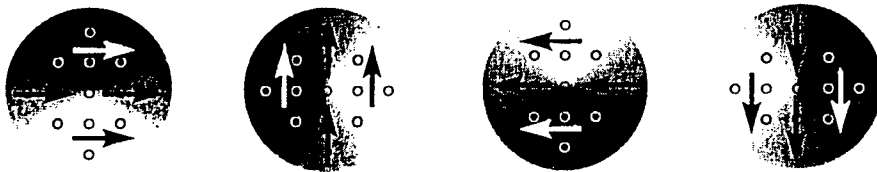
those that have been observed in the leech body wall mechanosensory system that controls a bending reflex (see Lewis and Kristan, 1998). In most tests, collator responses were recorded across a range of flow field positions relative to the receptive field of the collator neuron. For both uniform and gradient mappings, tests were also performed using a range of mapping "bandwidths". In the following equations for synaptic mappings, bw= the angular bandwidth of a uniform mapping, and "hbw" = the half bandwidth of a gradient mapping function.

Uniform synaptic weight function:

$$w = 1, \quad |\theta| <= bw/2 \quad\quad (6)$$
$$w = 0, \quad |\theta| > bw/2$$

Gradient synaptic weight function:

$$w = \cos(a*\theta), \quad |\theta| <= hbw \quad\quad (7)$$
$$w = 0, \quad\quad |\theta| > hbw$$

5-13

Results

Properties of Small-field motion detectors.

This section first describes the response profiles of the Reichardt-Hassenstein correlation-type sfmd models, then compares them with profiles obtained from the purely analytical sfmd model.

In both two-channel and three-channel versions of the correlation-type circuit (Fig. 4), maximal responses are exhibited at motion directions near zero degrees and motion speeds near approximately 120 deg/s. (The speed response maximum is determined by the value chosen for the internal delay). Note, however, that there is no single "preferred direction" that holds for all motion speeds. This is particularly evident in the two-channel model (Fig. 4A,B), in which the strongest responses form two distinct peaks, and responses at slow speeds show strong fluctuations as a function of motion direction. This complexity arises because the emd responses are determined by the temporal delay in the arrival times of a contrasting edge at different input channels, and this delay is affected not only by changes in motion direction, but also by changes in motion speed. At increasingly slow speeds, strong responses are still possible as long as the combination of motion direction and speed results in favorable delays. This phenomenon is related to that discussed by Zanker (1990) with reference to responses of *wide-field* motion-sensitive neurons to moving gratings.

The actual circuitry for sfmds of insects remains unknown, but almost certainly is more complex than the 2-channel correlator represented in Fig. 2A. Because neuronal morphologies and synaptic connections tend to be more diffuse, the more spatially distributed sfmd circuit of Fig. 2B may be more biologically realistic. This circuit also has the interesting property of

Fig. 4

A

C

B

D
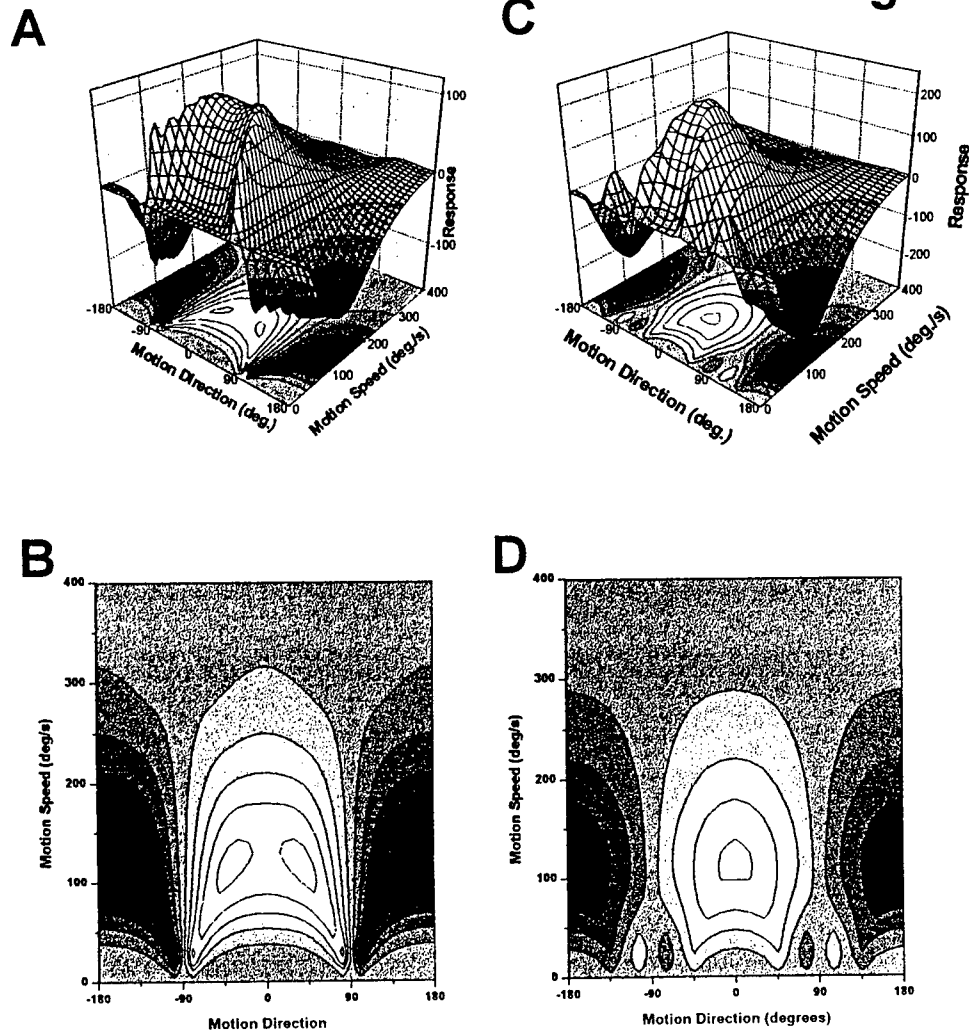
exhibiting reduced complexity in the response profile to motion direction and speed. As illustrated in Figs. 4C and D, the more distributed sfmd circuit produces only one major response maximum and one major minimum in the response profile. The shape of the profile is also more symmetrical and the contours less steep than in the 2-channel model. In summary,

computationally undesirable features of the simple two-channel model are reduced by assuming a more spatially distributed, and probably more biologically realistic version.

Figure 5 (F-I) shows the response profiles of the purely analytical model sfmds as a function of motion direction and speed, illustrated for the four alternative directional tuning functions f_A, f_B, f_C, f_D. The overall response profiles defined by the analytical model are qualitatively similar to those of the correlation-type sfmd models discussed above. As in the 3-channel version of the correlation model, each profile is characterized by a single response maximum that defines both a preferred direction and a preferred speed. As in versions of the correlation model that employ a subtraction stage, the analytical models that employ combined excitatory/inhibitory directional tuning (f_B and f_D) produce a response minimum at the preferred speed (when motion direction is 180 degrees from the preferred direction). The sfmds based on purely excitatory directional tuning (f_A and f_C) have a clearly defined response maximum, but no well-defined minimum.

II. Model Outputs: Optic Flow Selectivity in Wide-Field Collator neurons

Non-specialized (one-to-one) Synaptic Maps

The following two figures compare responses of the wide-field model to various optic flow field types and positions, also as a function of the four alternative sfmd directional tuning functions described in the Methods. First, figure 6 illustrates the properties of a simple circuit in which the spatial pattern of synaptic connection strengths from the sfmds to the collator neuron is completely uniform (one-to-one). With this unspecialized synaptic mapping pattern,

selectivity for optic flow *position* is obtained with three of the four directional tuning functions (f_A, f_C and f_D), but there is no selectivity for optic flow *type*. Clearly, whereas position selectivity is a simple consequence of pooling the outputs from an array of small-field detectors, optic flow type selectivity requires some level of nonuniformity in the pattern of synaptic weights. As subsequent results will demonstrate, however, it is possible to generate remarkably robust selectivity for optic flow type using any of a broad range of mappings.

Circuits for Flow Field Selectivity, I: Alternative Directional Tuning Functions

In order to test the properties of alternative directional tuning functions, a uniform synaptic mapping pattern was chosen based on a conceptual model of optic flow processing in primate cortex (Tanaka et al., 1989). A synaptic mapping pattern designed for clockwise rotation detection is illustrated in Figure 3 (upper patterns). Connections from sfmds whose "preferred direction" is rightward motion (0^0) are specified only from an upper quadrant of the sfmd array, connections from upward - sensitive (90^0) sfmds are made from the left quadrant of the array, and so on. Corresponding mapping patterns for counterclockwise rotation, expansion and contraction were specified simply by shifting the polar coordinates of the patterns for clockwise rotation by $+180^0$, -90^0, and $+90^0$, respectively. Similarly, patterns for non-cardinal optic flow types (such as clockwise rotation or counterclockwise expansion) were specified by using intermediate shifts.

Collator responses obtained using the 90^0 -bandwidth, uniform maps for pure clockwise rotation are shown in Figure 7, which has the same overall format as Figure 6. In contrast with the outputs from one-to-one mapping, all sfmd tuning functions now result in position selectivity

Fig. 5

A $f_A(\theta)$

B $f_B(\theta)$

C $f_C(\theta)$

D $f_D(\theta)$

E $f(S)$

Motion Speed (S)

Motion Direction (θ)

Response Amplitude

F $f_A(\theta,S)$

G $f_B(\theta,S)$

H $f_C(\theta,S)$

I $f_D(\theta,S)$

Response

Motion Direction

Motion Speed

5-18

Fig. 6

Fig. 7

A **Directional Function $F_A(\theta)$**

C **Directional Function $F_C(\theta)$**

B **$F_B(\theta)$**

D **$F_D(\theta)$**

Responses of Wide-field Collator

Stimulus Position

OPTIC FLOW STIMULI:
—— CLW Rotation
- - - - CCLW Rotation

—○— Expansion & Contraction

• Unidirectional

Receptive Field Diameter of Collator Neuron

to two or more of the cardinal flow types. More importantly, some degree of flow type selectivity is now obtained with all four directional tuning functions. Among the four tuning functions, clearly f_B (Fig. 7B) provides the most effective overall selectivity for clockwise rotation. Although other functions (Fig. 7A, C, D) also produce strong *responses* to centered clockwise rotation, sensitivity to unidirectional flow and complex responses to other flow types both compromise the desired *selectivity* for clockwise rotation, particularly in figure 7 A and C). Only with f_B are strong, position-sensitive responses to clockwise rotation combined with (1) symmetrical, inhibitory responses to the opposite flow type (counterclockwise rotation) and (2) complete insensitivity to contraction, expansion and unidirectional flow.

These results demonstrate that excellent selectivity for clockwise rotation can be obtained using a simple uniform synaptic mapping pattern designed to maximize responses to this cardinal flow type. Other tests have employed analogous uniform mappings designed to produce selectivity for flow types other than clockwise rotation (counterclockwise rotation, expansion, contraction, and intermediates). The results of these tests (not shown) are identical to those i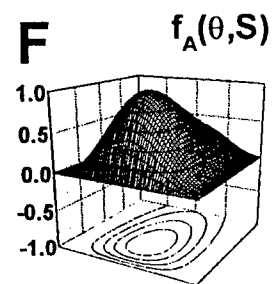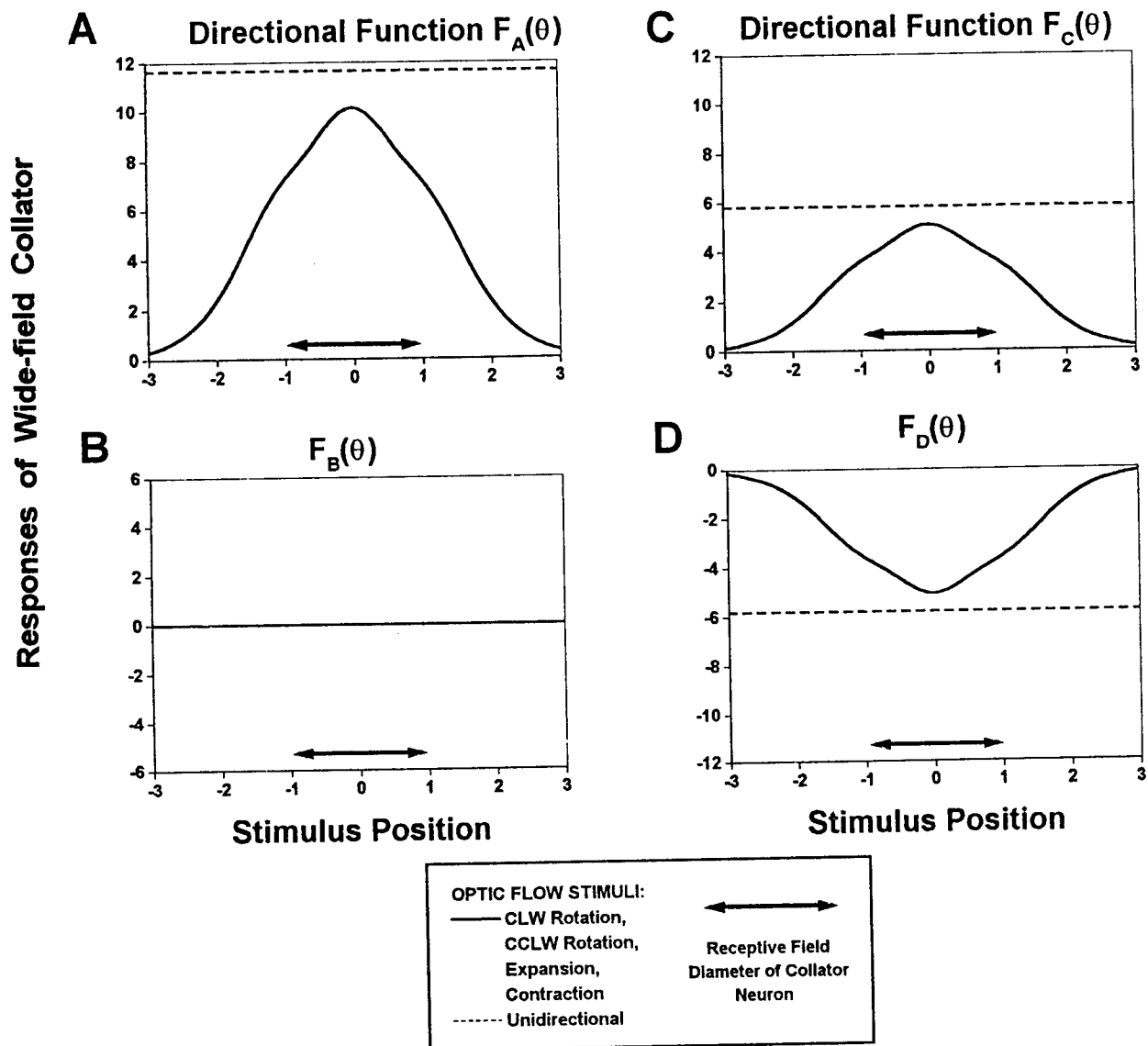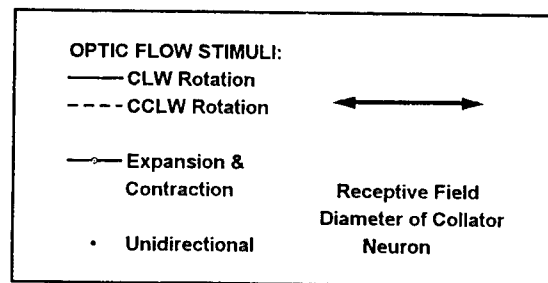n Figure 7, but with response patterns switched according to the flow type for which the mapping pattern was designed. For example, by using a uniform, 90^0-bandwidth mapping designed to detect pure expansion, f_B produces excitatory responses to expansion, inhibitory responses to contraction, and no response to unidirectional flow or to clockwise or counterclockwise rotation. In summary, excellent selectivity for any arbitrary flow type is possible by combining an appropriate mapping pattern with the broad, excitatory/inhibitory small-field directional tuning function f_B.

Robustness of optic flow selectivity at different motion speeds

In order to test further the robustness of selectivity for optic flow type in this model, responses using directional function f_B were obtained over a wide range of optic flow speeds, spanning the sfmd speed response function from zero to speeds far beyond the preferred local motion speed (as defined by the constant **k** in Equation 2). In response to stimulation with the "preferred" optic flow type defined by the synaptic maping pattern, position selectivity remains robust over a broad range of speeds, and is degraded only at the very slowest speeds (Fig. 8A). Moreover, sensitivity to non-preferred flow types (Fig. 8B) remains low across the entire range of motion speeds tested.

Coarse coding for optic flow type and position

A widespread feature of many neuronal information processing systems is that individual elements tend to be broadly tuned (coarsely coded) to each of several parameters. This important biological feature is reproduced quite well in this model. A broad range of optic flow fields, although they represent complex features of visual stimuli, can be represented on a linear, circularly symmetrical "flow type" scale that is analogous to simpler parameters such as angle and phase. In the example illustrated in Figure 9, the parametric values for pure expansion and contraction differ by 180^0, and rotational fields are orthogonal to looming fields. Conceivably, selectivity for flow field type could be either narrowly or broadly tuned. In this model, using sfmd function f_B, the tuning is broad indeed: the pattern of responses to optic flow type has a half-bandwidth of 180^0 and precisely fits a cosine function. Selectivity for flow field position is also coarsely-coded, as is demonstrated in Figure 9B for an expansion detector circuit.

Fig. 8
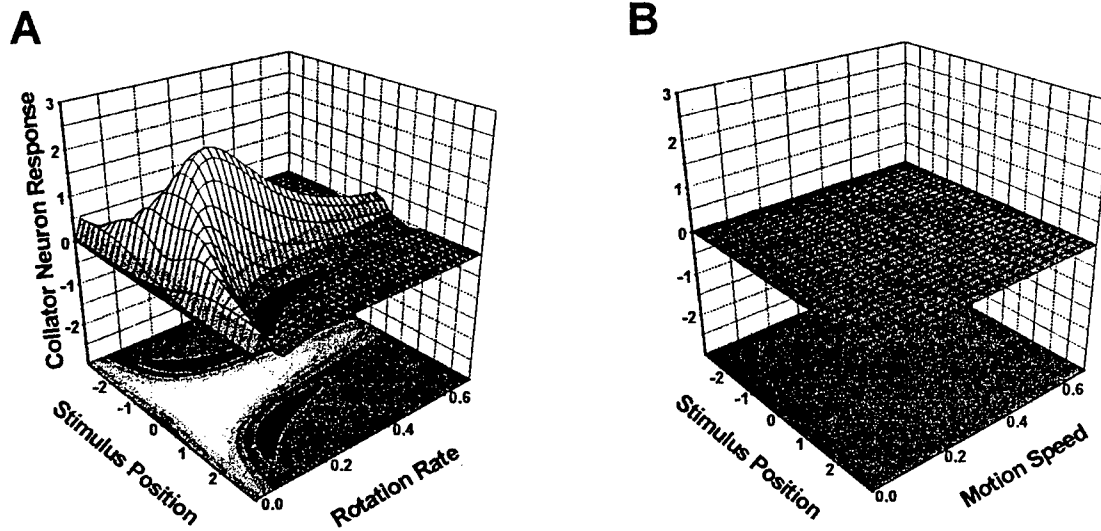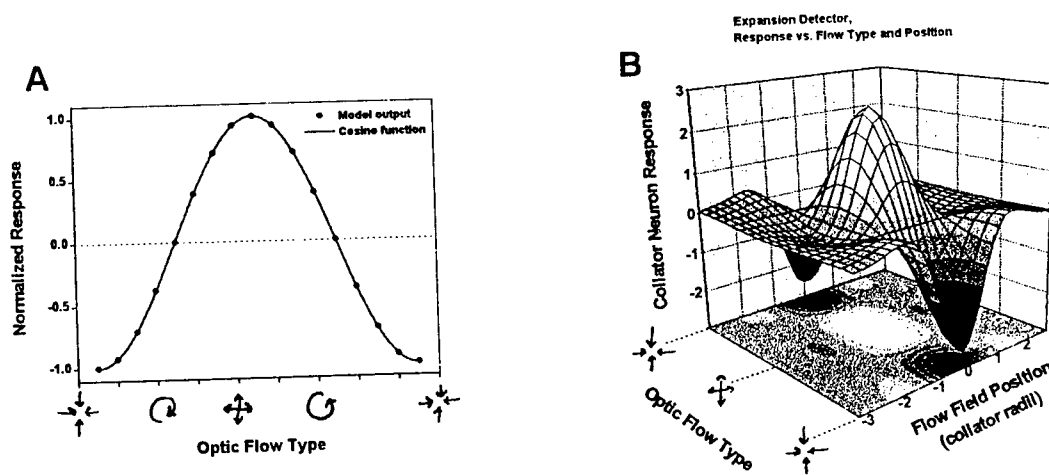
A

Collator Neuron Response — Stimulus Position — Rotation Rate

B

Stimulus Position — Motion Speed

Figure 9

A

Normalized Response vs. Optic Flow Type

- Model output
- Cosine function

B

Expansion Detector,
Response vs. Flow Type and Position

Collator Neuron Response — Optic Flow Type — Flow Field Position (collator radii)

The results described so far suggest that the circuit requirements for robust, coarsely-coded selectivity for optic flow type and position are remarkably simple: (1) a small-field processing level, composed of arrays of sfmds having (a) a suitable motion speed response function, (b) broad directional tuning, and (c) directional tuning that includes both excitatory and inhibitory responses (e.g. directional function f_B). (2) a wide-field "collation" stage, created by linearly summing the outputs of a mapped subset of the sfmds, the map for each sfmd array having been designed to promote maximal responses to a particular optic flow pattern when it is centered on the collator.

Despite their effectiveness, these specifications for an optic flow analyzer could suffer from at least two major shortcomings. First, it is possible that deviating from the 90^0 bandwidth, uniform maps employed so far could compromise the robust optic flow selectivity that has been observed. Second, the model appears to exhibit sensitive dependence upon suitable directional tuning of the small-field elements, suggesting that function f_B -like characteristics must be preserved at all times [in order to maintain effective optic flow processing] . If, however, certain details of directional tuning and output mappings are not crucial to the success of the circuit, far greater design flexibility would be possible. This kind of flexibility would be of considerable interest, both for understanding the evolutionary and developmental processes that have produced biological optic flow processing circuits, and also for evaluating the feasibility of manufacturing man-made processors based on similar designs. This section therefore investigates the extent to which alternative synaptic mappings may affect the model's

performance, and how to reduce the model's sensitivity to directional tuning characteristics through improvements to the synaptic mapping architecture.

In contrast with the uniform maps employed to this point, "gradient maps" are characterized by spatial gradients in synaptic connection strength (Fig. 3). Both types of mapping pattern are biologically plausible, and cosine-shaped gradient maps may be crucial to the performance of distributed-coding circuits that share features with this optic flow model (c.f. Lewis and Kristan, 1998). Thus, cosine-shaped gradient maps were tested as an alternative to uniform maps. Parts A-C of Figure 10 show the results of combining directional function f_B with cosine-shaped synaptic connectivity maps having half-bandwidths of 90, 180 and 270^0, respectively. As with the uniform maps presented above, the total range of synaptic weights is from 0 to 1. There is no qualitative change in either the position selectivity or optic flow type selectivity, and only minor changes in the maximum response amplitudes. Figure 10 D examines the relationship between maximum response amplitudes and map broadness (uniform map bandwidth or gradient map half-bandwidth) in more detail, and confirms that regardless of the type of mapping, strong response amplitudes are possible across a wide range of bandwidths.

The results in Figure 10 show that as long as the mapping bandwidth is not extremely narrow or broad, model performance is not highly sensitive to the precise bandwidth value. Having established that considerable design flexibility in mapping bandwidth is permissible, can adjustments of this parameter be used to reduce the model's sensitivity to sfmd directional tuning characteristics? As with the directional tuning functions, bandwidth is only one of two major characteristics of the mapping functions. The second, crucial feature is the offset from zero, which determines whether synaptic weights are dominated by excitatory or inhibitory

Fig. 10
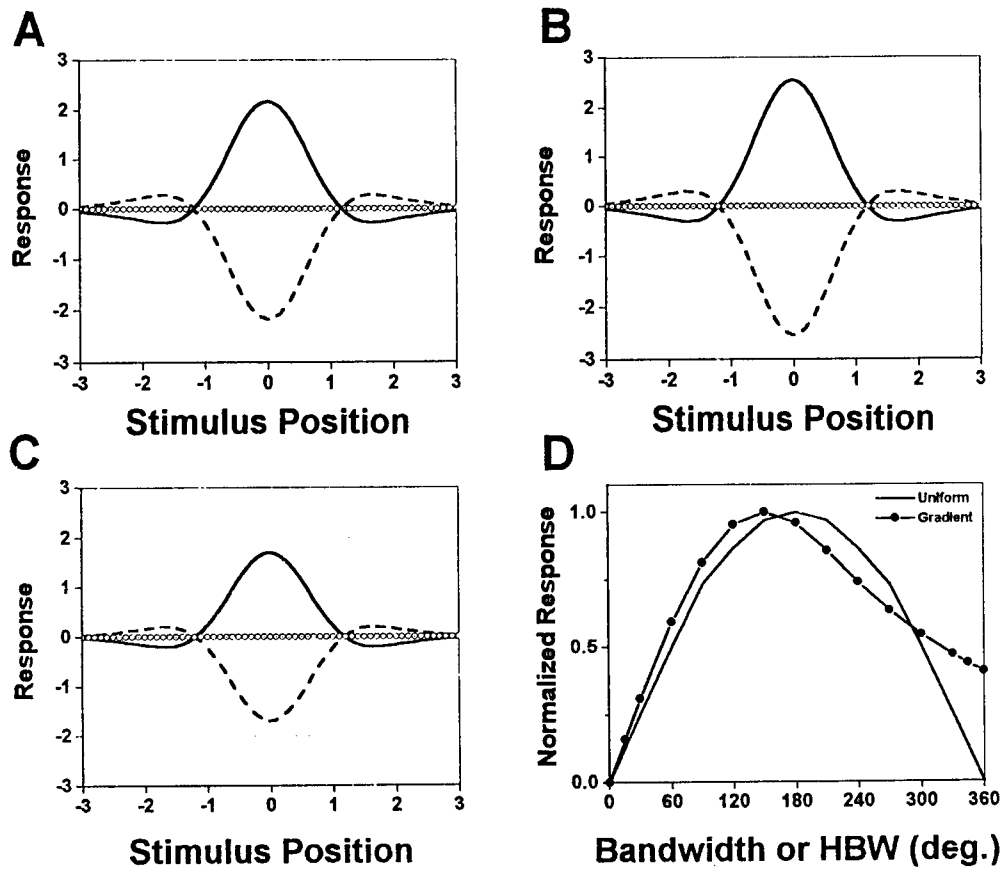
Fig. 11

A Directional Function $F_A(\theta)$

C Directional Function $F_C(\theta)$

B $F_B(\theta)$

D $F_D(\theta)$

Responses of Wide-field Collator

Stimulus Position

OPTIC FLOW STIMULI:

—— CLW Rotation
---- CCLW Rotation

—∘— Expansion & Contraction

• Unidirectional

←——→ Receptive Field Diameter of Collator Neuron
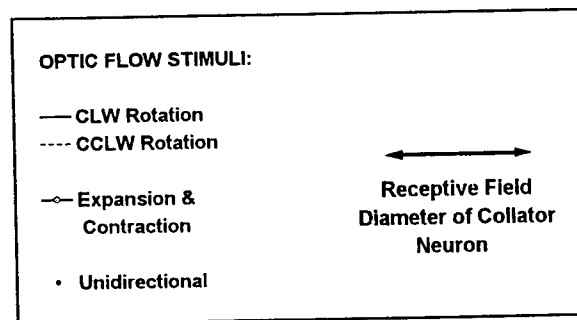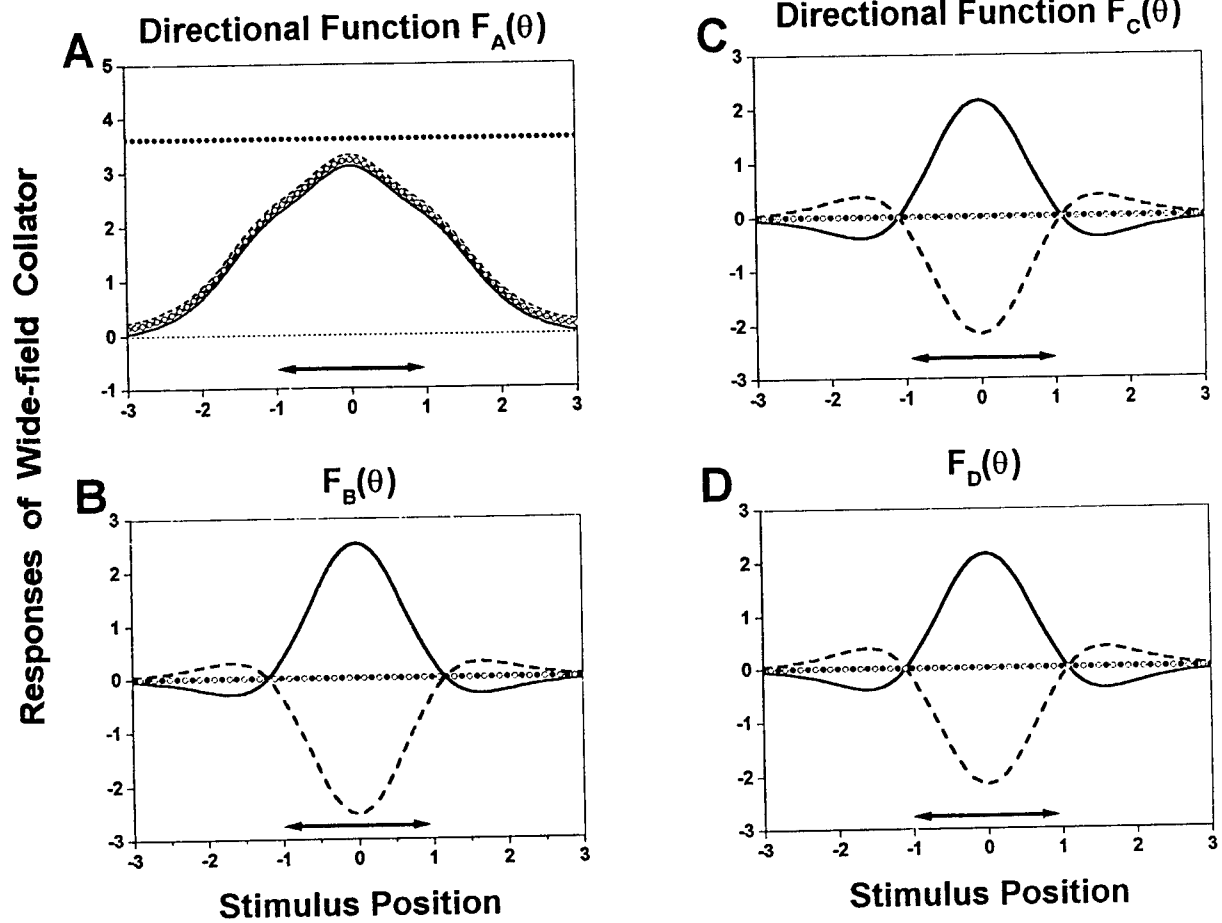
values. Preliminary tests indicate that an interplay between the mapping function offset and bandwidth determines the sensitivity to sfmd directional tuning. With cosine-shaped gradient mappings, a very effective (and possibly optimal) solution employs a half-bandwidth of 180^0 and a zero offset. As illustrated in Figure 11, this mapping function nearly eliminates the model's dependence on sfmd tuning characteristics. Now, functions f_B, f_C, and f_D all produce excellent selectivity for flow field type and position; only f_A remains ineffective.

Discussion

This computational model combines biologically realistic responses of small-field motion detectors to motion direction and speed, with hypothesized spatial and physiological features of synaptic connections from the small-field detectors to a wide-field, linear summation stage. These simple design features were inspired by known structural and functional characteristics of motion-processing circuits in insect brains. Extensive tests of the model's responses to various flow field stimuli demonstrate robust selectivity for flow type and position. Moreover, these tests demonstrate considerable design flexibility in the specification of (a) directional tuning functions of the small-field motion detectors, and (b) spatial patterns of output connections to the collator stage. The model makes various predictions regarding the functional organization of optic flow processing circuitry in insects, many of which can be tested using existing technology for neurophysiological and neuroanatomical investigations. In addition, the model suggests which design features will be crucial for future biomimetic machine vision hardware. By analogy, the same features should also be useful for other flow field applications, such as hydrodynamic flow analysis using mechanosensors instead of photosensors. Because of the

robust behavior and flexible design features of this model, it can provide a strong basis for future designs of inexpensive, yet sophisticated guidance control systems for munitions and other automated vehicles. The main conclusions from this investigation are discussed below.

This model makes various predictions, several of which are testable with physiological recordings and anatomical studies of insect brains.

1. The well-known 4-layered directional organization of the calliphorid fly lobula plate (Strausfeld, 1989) is not required for optic flow processing. This leads to a hypothesis: not all dipterans need to have the 4-layered lobula plate structure, and in those that do, this must have some other, as yet unidentified functional significance. Current research is also testing the assumption that the sfmd outputs (T5 cells) are indeed segregated to 4 distinct strata in the lobula plate.

2. Directional tuning properties of sfmds, and their synaptic mappings to optic flow-selective wide-field neurons are interrelated, such that any physiological data regarding one type of property will suggest predictions about the other. For example, if directional tuning of sfmds varies according to stimulus conditions, the model predicts that optic flow analyzers should have sophisticated synaptic mapping patterns (e.g. involving 2 or more neurotransmitters). Conversely, if spatial synaptic maps involve purely excitatory transmitters, directional tuning functions are expected to be similar to function fB (Fig. 5). In this case, the model predicts that additional mechanisms are required to stabilize sfmd directional tuning functions under varying input conditions (e.g. different spatiotemporal frequencies). Additional recordings from sfmds

are needed in order to test their response profiles as a function of both motion direction and speed.

3. The model predicts "coarse coding" for flow field type, as has been observed in primates (Graziano et al. 1994). This prediction can be tested in real optic flow neurons in insects. The model also predicts selectivity for flow field position, with details dependent on the speed tuning of the sfmds relative to the receptive field size of the collator neurons.

4. The model makes detailed predictions of the spatial patterns of excitatory and inhibitory inputs to dendrites (or postsynaptic spatial patterns of dendritic activity) of collator neurons (JK Douglass, in preparation). In principle, these predictions are testable with intracellular recordings using voltage-sensitive or calcium-sensitive dyes.

What are the key design features for optic flow processing in this model? Optic flow-selective neurons with realistic biological properties can be made using a simple feed-forward network composed of two basic stages: (a) a few arrays (here, four) of correlation-type sfmds that respond according to temporal delays determined by motion direction and speed, and (b) simple spatial patterns of purely excitatory synaptic connection strengths that can be designed to generate selectivity for a wide variety of flow fields. Thus, a single set of sfmd arrays can be used to generate model collator neurons that are selective for rotation, expansion, contraction, unidirectional flow, or intermediate flow types. This result provides a detailed demonstration of a basic principle of information processing that has been documented in insect visual systems: shared, evolutionarily conserved peripheral mechanisms (Douglass and Strausfeld 1995; Buschbeck and Strausfeld, 1996) provide the basis for various types of higher-level processing (Buschbeck and Strausfeld, 1997).

In summary, the biomimetic flow-selective properties exhibited by this model include "coarse coding" for optic flow type, maintenance of selectivity for optic flow type across a range of optic flow speeds, and sensitivity to optic flow position relative to receptive field center of a flow-selective collator neuron or summating element. These selectivities can be produced by optimizing either of two basic design features: the response properties of the sfmds and the spatial synaptic output mappings to the collator stage. If roughly optimal values for either of these two features are met, robust model behavior can be had without the need for tight control of the other parameters. The obvious implication for circuit design is that by identifying which design parameters can be optimized more precisely and at lower production cost, tremendous savings can be realized because of the increased flexibility in achieving those parameters that would be more expensive to specify according to stricter tolerances.

References

Borst A, Egelhaaf M (1989) Principles of visual motion detection. Trends Neurosci 12(8): 297-306.

Borst A and M Egelhaaf, 1990. Direction selectivity of blowfly motion-sensitive neurons is computed in a two-stage process. Proc. Natl. Acad. Sci. USA 87: 9363-9367.

Buchner E, 1976. Elementary movement detectors in an insect visual system. Biol Cybern 24, 85-101.

Buchner, E., S. Buchner and I. Bülthoff. (1984) Deoxyglucose mapping of nervous activity induced in *Drosophila* brain by visual movement. J. Comp. Physiol. A *155*:471-483.

Buschbeck EK, Strausfeld NJ, 1996. Visual motion detection circuits in flies: Small-field retinotopic elements responding to motion are evolutionarily conserved across taxa. J Neurosci 16, 4563-4578.

Buschbeck EK, Strausfeld NJ, 1997. The relevance of neural architecture to visual performance: Phylogenetic conservation and variation in dipteran visual systems. J Comp Neurol 383, 282-304.

Douglass JK 1997. Guiding missiles "On the fly:" Applications of neurobiological principles to machine vision for armaments. Final Report for *Summer Faculty Research Program, Air Force Office of Scientific Research*. pp. 20-1 to 20-20.

Douglass JK and Strausfeld NJ, 1995. Visual motion detection circuits in flies: Peripheral motion computation by identified small-field retinotopic neurons. J Neurosci 15, 5596-5611.

Douglass JK, Strausfeld NJ, 1996. Visual motion-detection circuits in flies: Parallel direction- and non-direction-sensitive pathways between the medulla and lobula plate. J Neurosci 16, 4551-4562.

Douglass JK and Strausfeld, NJ, 1998. Functionally and anatomically segregated visual pathways in the lobula complex of a calliphorid fly. *J Comp Neurol* 396, 84-104.

Duffy CJ and Wurtz RH, 1991. Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli. J Neurophysiol 65:1346-1359.

Eckert H, 1980. Functional properties of the H1-neuron in the third optic ganglion of the blowfly, *Phaenicia*. J comp Physiol 135, 29-39.

Eckert H, 1982. The vertical-horizontal neurone (VH) in the lobula plate of the blowfly, *Phaenicia.* J Comp. Physiol. 149, 195-205.

Fischbach K-F, Dittrich APM (1989) The optic lobe of *Drosophila melanogaster.* I. A Golgi analysis of wild-type structure. Cell Tissue Res 258:441-475.

Graziano MSA, Andersen RA and Snowden RJ, 1994. Tuning of MST neurons to spiral motions. J Neurosci 14:54-67.

Grzywacz, N., E. Sernagor and F. R. Amthor. (1995) Directional selectivity in the retina. In M.A. Arbib (ed): The handbook of brain theory and neural networks. Cambridge, MA, M.I.T. Press, pp. 312-314.

Hausen, K, 1982. Motion sensitive interneurons in the optomotor system of the fly. II. The Horizontal cells: Receptive field organization and response characteristics. Biol Cybern 46, 67-79.

Hengstenberg R, 1982. Common visual response properties of giant vertical cells in the lobula plate of the blowfly *Calliphora erythrocephala.* J. Comp. Physiol. A 149, 179-193.

Koenderink JJ, 1986. Optic flow. Vision Res 26:161-180.

Krapp, H. G. and R. Hengstenberg. (1996) Estimation of self-motion by optic flow processing in single visual interneurons. Nature *384*:463-466.

Krapp, H. G. and R. Hengstenberg. (1997) A fast stimulus procedure to determine local receptive field properties of motion-sensitive visual interneurons. Vision Res. *37*:225-234.

Laughlin S (1981) Neural principles in the peripheral visual systems of invertebrates. In: Comparative physiology and evolution of vision in invertebrates. Handbook of Sensory Physiology VII/6B (Autrum H, ed), pp 133-280. Berlin: Springer.

Lewis JE and WB Kristan, Jr., 1998. A neuronal network for computing population vectors in the leech. Nature 391:76-79.

Maunsell JHR, Van Essen DC, 1983. Functional properties of neurons in middle temporal visual area of the macaque monkey. I: Selectivity for stimulus direction, speed, and orientation. J Neurophysiol 49:1127-1147.

Morrone MC, Burr DC and Vaina LM, 1995. Two stages of visual processing for radial and circular motion. Nature 376:507-509.

Strausfeld, N. J. (1989) Beneath the compound eye: Neuroanatomical analysis and physiological correlates in the study of insect vision. In: Facets of vision (Stavenga DG, Hardie RC, eds), pp 317-359. Heidelberg: Springer.

Strausfeld, N. J. and J.-K. Lee (1991) Neuronal basis for parallel visual processing in the fly. Visual Neurosci. 7:13-33.

Tanaka K, Fukada Y, Saito H, 1989. Underlying mechanisms of the response specificity of expansion/contraction and rotation cells in the dorsal part of the medial superior temporal area of the macaque monkey. J Neurophysiol 62:642-656.

Theunissen FE and Miller JP, 1991. Representation of sensory information in the cricket cercal sensory system. II. Information theoretic calculation of system accuracy and optimal tuning-curve widths of four primary interneurons. J Neurophysiol 66, 1690-1703.

Zanker JM, 1990. On the directional sensitivity of motion detectors. Biol. Cybern. 62:177-183.

PREDICTION OF COMPRESSION TEXTURESS IN TANTALUM BY A PENCIL-GLIDE COMPUTER PROGRAM

Dr. William Hosford
Associate Professor
Department of Materials and Science

University of Michigan College of Engineering
2300 Hayward Street
Ann Arbor, Michigan 48109-2136

Final Report on AFSOR Contract F49620-93-C-0063
**Prediction of Compression textures in Tantalum
by a Pencil-Glide Computer Program.**

Background

For bcc metals deforming by <111>-pencil glide, lattice rotations can be uniquely predicted. The purpose of this project is to predict the texture development in tantalum during compression. After heavy compression, of bcc metals, the grains become oriented so that either their <111> directions or <100> directions are aligned with the axis of compression (1).

An existing computer program (2) based on a true <111>-pencil glide model has been used to relate the plastic properties of bcc metals to their texture. The present work is to modify this program so that it can predict orientation changes during axisymetric compression. This program recognizes that with <111>- pencil glide, there are eight degrees of freedom; these degrees of freedom being the amount of slip in each of the four <111> slip directions and the orientation of the slip planes associated with each of the four directions. For any general shape change, five degrees of freedom of slip are necessary, so a general shape change can be accomplished either by slip in three <111> directions or by slip on all four <111> directions, with the orientations of the slip planes providing the extra degrees of freedom.

The calculation procedure explores both possibilities. For the three active slip direction possibility, one slip direction is assumed to be inactive. Then an orientation of one of the slip planes is assumed. The orientations of the other two are then calculated, as well as the amount of slip in all three directions, the total amount of slip, $\Sigma d|\gamma_i|/d\varepsilon$, being noted. Then a new orientation of the slip plane is assumed and the calculation is repeated. Such calculations are continued until the minimum value of $\Sigma d|\gamma_i|/d\varepsilon$ is found. This procedure is then repeated, each time with a different inactive slip direction. Finally, the solution with the lowest value of $\Sigma d|\gamma_i|/d\varepsilon$ is noted.

The other possibility is that all four slip directions are active. Piehler and Backofen (3) found that there are only four stress states that can activate all four slip directions simultaneously. The virtual work to achieve the shape change is calulated for each. The best of the four-slip direction solutions is the one for which the virtual work is a maximum. The optimum four slip system solution is then examined to see if the stress state required to produce simultaneous slip on the four systems is admissible. If it is, this is the appropriate solution. Otherwise the three

slip system solution is appropriate.

Dr. William Hosford

The necessary work is to identify the slip systems operative for the appropriate solution and the amount of slip on these and use this information to calculate the lattice rotations. The mathematics required for the analysis has been worked out and is outlined below.

Orientation description

The system used in this project of describing the orientation of an axis, z, relative to the cubic crystal axes involves two angles, θ and ϕ as shown ion figure 1. The angle ϕ can be thought of as the latitude and θ as the longitude if 3 = [001] is the north pole and [100] lies on the equator at the Greenwich longitude.
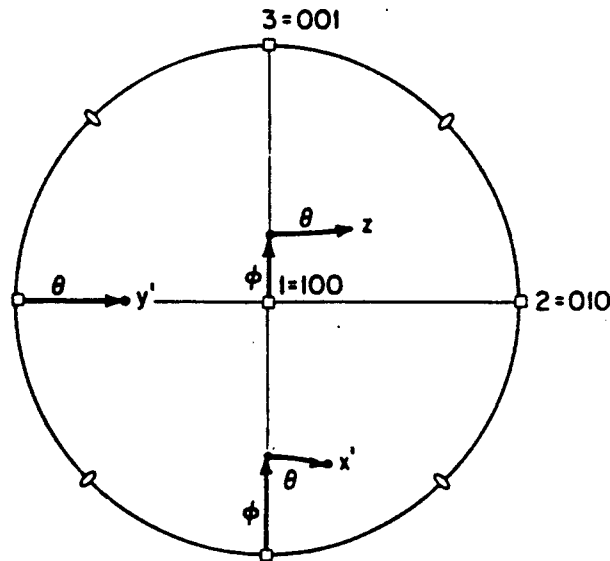

Figure 1. Stereographic representation of the orientation of the z-axis, in terms of θ and ϕ.

Shape change.

The objective of the program is to determine the orientation changes of grains of various orientations undergoing deformation that is axially symetric about z. Therefore the input shape changes will be in small increments of compressive strain. (e.g. $\Delta\varepsilon_z$ = -0.1). For axisymetry then

$$\Delta\varepsilon_x = \Delta\varepsilon_y = - (1/2)\Delta\varepsilon_z \tag{9}$$

These strains can be transformed into strains along the cube axes by the following equations:

$$\Delta\varepsilon_1 = \Delta\varepsilon_z[\cos^2\phi\cos^2\theta - .5(\sin^2\phi\cos^2\theta + \sin^2\theta)] \tag{10}$$

$$\Delta\varepsilon_2 = \Delta\varepsilon_z[\cos^2\phi\sin^2\theta - .5(\sin^2\phi\sin^2\theta + \cos^2\theta)] \tag{11}$$

$$\Delta\varepsilon_3 = \Delta\varepsilon_z[\sin^2\phi - .5\cos^2\phi] \tag{12}$$

$$\Delta\gamma_{23} = \Delta\varepsilon_z[2\sin\phi\cos\phi \sin\theta + \cos\phi\sin\phi\sin\theta] \tag{13}$$

$$\Delta\gamma_{31} = \Delta\varepsilon_z[2\sin\phi\cos\phi \cos\theta + \cos\phi\sin\phi\cos\theta] \tag{14}$$

$$\Delta\gamma_{12} = \Delta\varepsilon_z3\sin\theta\cos\theta3\cos^2\phi \tag{15}$$

January 26, 1999

In the <111>- pencil glide model, it is assumed that the slip directions are the close-packed <111> directions and the slip planes are those planes containing for which the shear stress is the highest. Figure 2 shows how the orientations of the normals to these planes can be described by an angle, ψ, of rotation about the slip directions. The direction cosines between these slip elements and the 1 = [100], 2 = [010] and 3 = [001] crystal axes are listed in Table I.

Figure 2. Stereographic representation of the pencil glide systems. a, b, c, and d are the slip plane normals for the A = $[1\bar{1}1]$, B =[111], C = $[11\bar{1}]$ and D = $[1\bar{1}\bar{1}]$ slip directions. ψ_a, ψ_b, ψ_c, and ψ_d, are the rotations of the slip plane normal about these axes.
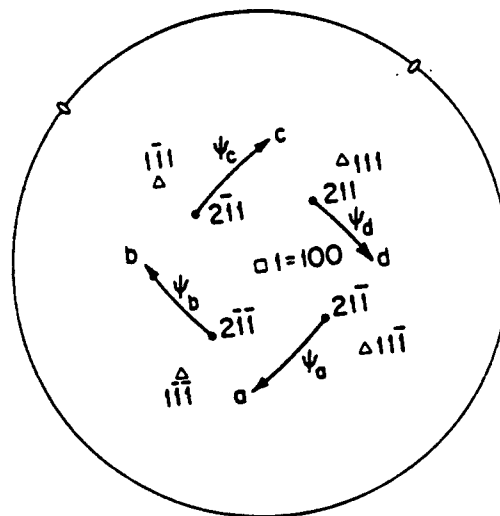
Table I   Direction cosines of the <111> slip directions and slip-plane normals with the cubic axes 1 = [100], 2 = [010], and 3 = [001]

Slip directions	slip plane normals
A = $[1\bar{1}1]$	
$\ell_{1d} = \sqrt{(2/3)}\cos\psi_a$	$\ell_{1n} = 1/\sqrt{3}$
$\ell_{2d} = \sqrt{(2/3)}\cos(60+\psi_a)$	$\ell_{1n} = -1/\sqrt{3}$
$\ell_{3d} = -\sqrt{(2/3)}\cos(60-\psi_a)$	$\ell_{1n} = 1/\sqrt{3}$
B = [111]	
$\ell_{1d} = \sqrt{(2/3)}\cos\psi_b$	$\ell_{1n} = 1/\sqrt{3}$
$\ell_{2d} = -\sqrt{(2/3)}\cos(60-\psi_b)$	$\ell_{1n} = 1/\sqrt{3}$
$\ell_{3d} = \sqrt{(2/3)}\cos(120-\psi_b)$	$\ell_{1n} = 1/\sqrt{3}$
C = $[11\bar{1}]$	
$\ell_{1d} = \sqrt{(2/3)}\cos\psi_c$	$\ell_{1n} = 1/\sqrt{3}$
$\ell_{2d} = \sqrt{(2/3)}\cos(120-\psi_c)$	$\ell_{2n} = 1/\sqrt{3}$
$\ell_{3d} = -\sqrt{(2/3)}\cos(60-\psi_c)$	$\ell_{3n} = -1/\sqrt{3}$
D = $[1\bar{1}\bar{1}]$	
$\ell_{1d} = \sqrt{(2/3)}\cos\psi_d$	$\ell_{1n} = 1/\sqrt{3}$
$\ell_{2d} = -\sqrt{(2/3)}\cos(60-\psi_d)$	$\ell_{2n} = -1/\sqrt{3}$
$\ell_{3d} = \sqrt{(2/3)}\cos(60+\psi_d)$ $\ell_{3n} = -1/\sqrt{3}$	

January 26, 1999

Rotations

Slip on each of the slip systems results in a rotation, ω_i, of the crystal orientation about the three crystal axes, that depends on the orientation of the slip plane, ψ_i, and the amount of slip, γ_i, on that system. The net rotations about each of the crystal axes are the sums of the rotations, ω_i, cause be each slip system. The equations relating the rotations about the cubic axes to the individual values of ψ_i, and γ_i are listed below. For rotation about the 1 = [100] axis,

$\omega_1 = \Sigma(\gamma/2)(\ell_{2D}\ell_{3N} - \ell_{2N}\ell_{3D})$ summed over all 4 slip systems or

$$\omega_1 = (\sqrt{3}/6)[\gamma_a(\ell_{2Da} + \ell_{3Da}) + \gamma_b(\ell_{2Db} - \ell_{3Db}) + \gamma_c(-\ell_{2Dc} - \ell_{3Dc}) + \gamma_d(-\ell_{2Dd} + \ell_{3Dd})]$$

(16)

For rotation about the 2 = [010] axis,

$\omega_2 = \Sigma(\gamma/2)(\ell_{3D}\ell_{1N} - \ell_{3N}\ell_{1D})$ summed over all 4 slip systems or

$$\omega_2 = (\sqrt{3}/6)[\gamma_a(\ell_{3Da} - \ell_{1Da}) + \gamma_b(\ell_{3Db} - \ell_{1Db}) + \gamma_c(\ell_{3Dc} + \ell_{1Dc}) + \gamma_d(\ell_{3Dd} + \ell_{1Dd})]$$

(17)

For rotation about the 3 = [001] axis,

$\omega_3 = \Sigma(\gamma/2)(\ell_{1D}\ell_{2N} - \ell_{1N}\ell_{2D})$ summed over all 4 slip systems or

$$\omega_3 = (\sqrt{3}/6)[\gamma_a(-\ell_{1Da} - \ell_{2Da}) + \gamma_b(\ell_{1Db} - \ell_{2Db}) + \gamma_c(\ell_{1Dc} - \ell_{2Dc}) + \gamma_d(-\ell_{1Dd} - \ell_{3Dd})]$$

(18)

Expressing rotations about cube axes as changes of ϕ and θ.

The change $\Delta\phi$ resulting from a rotation ω_1, can be deduced from the spherical triangles represented in figure 3 as follows:

Figure 3. Change in ϕ and θ from a rotation ω_1

$\sin\phi = \cos(90-\phi) = \sin\rho\cos\psi_1$, so

$\psi = \sin\phi/\sin\rho_1$.

(19)
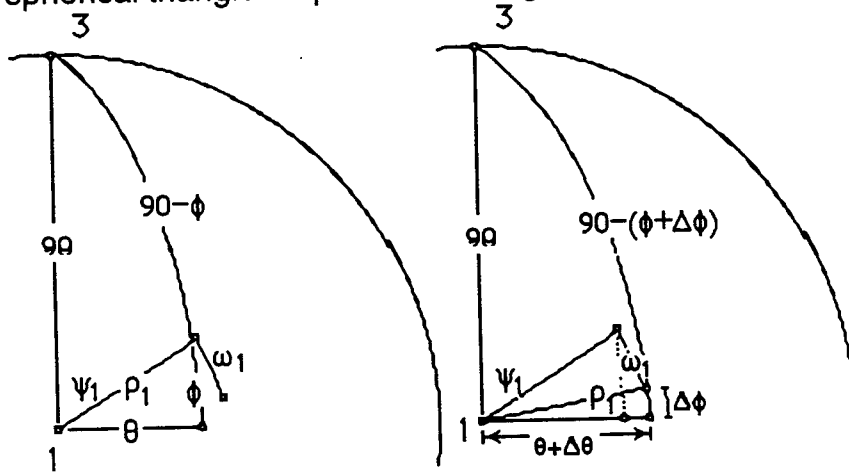
Here $\cos\rho_1 = \cos\theta\cos\phi$ is a constant during rotation, ω_1.

Since $\sin[\phi + (\Delta\phi)_1] = \sin\rho_1\cos(\psi_1 + \omega_1)$, $\cos[\phi + (\Delta\phi)_1] = \cos\rho_1\cos[\theta + (\Delta\theta)_1]$, so

January 26, 1999   6-5

the change in q can be found from

$$\cos[\theta+(\Delta\theta)_1] = \cos[\phi+(\Delta\phi)_1]/\cos\rho_1 \tag{20}$$

The change $\Delta\phi$ resulting from a rotation ω_2, can be found from the spherical triangles represented in figure 4.

Figure 4. Change in ϕ and θ from a rotation ω_2

$$\sin[\phi+(\Delta\phi)_2] = \sin\rho_2\cos(\psi_2 +\omega_2), \tag{21}$$

where $\cos\rho_2 = \sin\pi\sin\theta$ and $\cos\psi_2 = \sin\phi\sin\rho_2$ so

$$\sin[\theta +(\Delta\theta)_2] = \sin\phi/\sin\rho_2. \tag{22}$$

The change $\Delta\phi$ resulting from a rotation ω_3, can be found from the spherical triangles represented in figure 5.

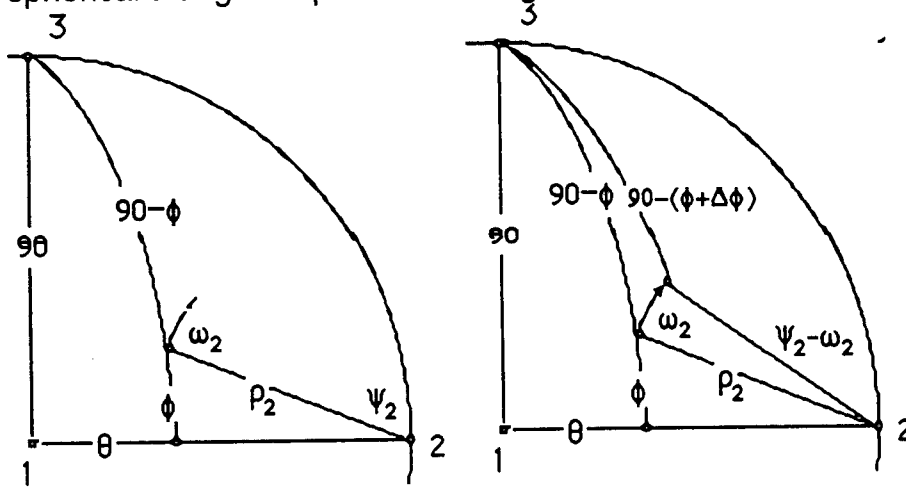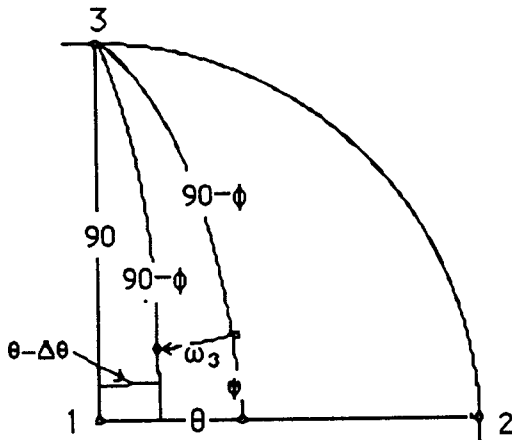Figure 5. Changes in ϕ and θ from a rotation ω_2

$\rho_3 = 90 - \phi$ so

$$\omega_3 = -(\Delta\theta)_3 \tag{23}$$

$$(\Delta\phi)_3 = 0 \tag{24}$$

6-6

January 26, 1999

The overall orientation change is then given by   (25)

$$\Delta\phi = (\Delta\phi)_1 + (\Delta\phi)_2$$

and   (26)

$$\Delta\theta = (\Delta\theta)_1 + (\Delta\theta)_2 + (\Delta\theta)_3$$

Overall analysis

By applying sucessive increments of $\Delta\varepsilon_z$, the changes of orientation with compressive strain will be monitored. The results will allow us to predict which orientations will rotate toward an [100]-end orientation and which toward a [111]-end orientation.

The results will allow prediction of the strain level necessary to develope a given strength of texture. The validity of this approach will be checked by comparing these predictions with experimental observations.

Work to date

Work on this project is far behind schedule. Money for the project was not received by the University of Michigan until May 1998. Because of this it was not possible to hire a student to work on the project until January 1999 when Daniel Greszczak was hired.

The existing Fortran program for pencil glide described above has been acquired. Modififications to this program for axisymetric deformation about a single crystallographic direction are being made. This has involved becoming familiar with the program and changing of the notation for axes.

The necessary mathematical equations (1 to 26) for calculating lattice rotations from slip have been worked out.

1.   C. S. Barrett, *Trans. AIME,* **105** (1939) p. 296.
2.   R. W. Logan, "Upper-Bound Anisotropic Yield Loci Calculations Assuming <111>-Pencil Glide," *Int. J. Mech. Sci.* **22** (1980) 419-59.
3.   H. R. Piehler and W. A. Backofen, "A Theoretical Examination of the Plastic Properties of BCC Crystals Deforming by <111> Pencil Glide," *Met. Trans.* **2** (1971) 249-55.

January 26, 1999

Parallel Implementation of Computational Electromagnetics Simulation Using High Performance Fortran

Yi Pan
Associate Professor
Department of Computer Science

The University of Dayton
300 College Park
Dayton, OH 45469-2160

Parallel Implementation of Computational Electromagnetics Simulation Using High Performance Fortran

Yi Pan
Associate Professor
Department of Computer Science
The University of Dayton

Abstract

An efficient and scalable parallelization of a sequential Fortran time-dependent Maxwell equations solver using High Performance Fortran (HPF) has been implemented. This report describes the background of the project, the theory behind the efficiency being achieved, the parallelization methodologies employed, and the experimental results obtained on the Cray T3E massively parallel computing system.

The program is parallelized through HPF directives. The major directives used in the code are the ALIGN, DISTRIBUTE, and INDEPENDENT directives. Through these directives, we were able to distribute the arrays according to our parallelization strategy and to instruct the parallel compiler how to parallelize the code efficiently.

Experimental runs show that the execution time is reduced drastically through parallel computing. Due to the requirement of the HPF compiler I used, many huge arrays have to be initialized before their usage, a lot of overhead is introduced. However, the code is still scalable up to 98 processors on the Cray T3E. In fact, based on our experience we predict that the execution time can be further reduced by using 196 processors. Unfortunately, the T3E system we used has only 128 processors and we cannot confirm this.

Based on the experimentation carried out in this research, we believe that a high level parallel programming language such as the High Performance Fortran is a fast, viable and economical approach to parallelize many existing sequential codes which exhibit a lot of parallelism.

Parallel Implementation of Computational Electromagnetics Simulation Using High Performance Fortran

Yi Pan

1 Introduction

Computational electromagnetics (CEM) in the time domain is the most general numerical approach for describing dynamic or wide-band frequency electromagnetic phenomena. Computational simulations are derived from discretized approximations to the time-dependent Maxwell equations [1-10]. High numerical efficiency of CEM simulation procedures can be attained either by algorithmic improvements to solve the Maxwell equations or by using scalable parallel distributed memory computer systems. Since the massive volume of data processing are involved in solving the Maxwell equations, distributed memory computer systems are viable means to solve the memory shortage problem on workstations or vector computer systems. The other advantage is reduced time when parallel processing is employed to solve the Maxwell equations. Hence, parallelization of existing sequential Fortran code for solving Maxwell equations is an important effort towards developing efficient and accurate CEM code in analyzing refraction and diffraction phenomena for aircraft signature technology.

Previously, three versions of of the solver were available. One is the sequential Fortran version developed by Dr Joseph J.S. Shang for vector machines such as the CRAY C90. The second is the MPI (Message Passing Interface) [15-16] implemented by Dr. Marcus Wagner of IBM. The third one is the Power Fortran version [14] developed by me during the summer of 1997 when I was an AFOSR Faculty Research Fellow at WPAFB. The description of the programs and their performance results are summarized in [11]. Using the message passing paradigm has several disadvantages: the cost of producing a message passing code may be between 5 and 10 times that of its serial counterpart, the length of the code grows significantly, and it is much less readable and less maintainable than the sequential version. For these reasons, it is widely agreed that a higher level programming paradigm is essential if parallel systems are to be widely adopted. The results obtained during my summer research show that the performance of the parallel PFA code is even better than the MPI counterpart besides the advantages stated above. This suggests that a higher level programming paradigm such as HPF is a viable and alternative way to parallelize CEM codes.

However, the results achieved during my summer research is only a preliminary step in achieving a scalable version of the MAX3D code. The major problem with the PFA version is that it is not portable across different parallel computer systems since Power Fortran is only

available on SGI Power Challenger and Origin 2000.

The primary research objectives of this research are to develop parallel code for the MAX3D using High Performance Fortran (HPF). HPF is more powerful than Power Fortran and is available in almost every major parallel computer systems. Our goals are to further improve the scalability of the Power Fortran version of the MAX3D code, investigate the possibility of using HPF to implement parallel CEM applications, make the parallel code portable across several parallel platforms, and present the results to other researchers via conference and/or journal publication. These objectives have been accomplished through implementation of the parallel code for the MAX3D via PGHPF (a HPF compiler developed by Portland Group). Experimental runs show that the execution time is reduced drastically through the usage of HPF. The time spent for the implementation is about two months. Compared with MPI implementation, it is much shorter. Since HPF is a general parallel language and not targeted to any particular machine, its performance is usually not as good as specialized parallel language such as Power Fortran on the SGI machines. Yet, the HPF code we produced is still scalable up to 98 processors on the Cray T3E. Based on the experimentation carried out in this research, we show that the High Performance Fortran is a viable and alternative way to parallelize many existing sequential codes which exhibit a lot of parallelism besides MPI. In many cases, using HPF to implement a sequential Fortran code is more economical than using MPI. HPF also solves the problem of limited memory size on a single processor and provide a fast way to port the code on parallel computer systems.

In the following sections, I will present the work performed, the methods used, and the results achieved during the research, and outline the impact of the proposed research.

2 Methodologies

High Performance Fortran (HPF) is an informal standard for extensions to Fortran to assist its implementation on parallel architectures, particularly for data-parallel computation [18]. Among other things, it includes directives for expressing data distribution across multiple memories, extra facilities for expressing data parallel and concurrent execution, and a mechanism for interfacing HPF to other languages and programming models. HPF is available on almost any popular parallel computer systems, such as T3E, SP2, SGI Origin 2000, and is much easy to use and fast to port than MPI [15-16]. HPF allows you to add HPF directives in the sequential Fortran code and to apply the capabilities of a multiprocessor system to the execution of a single job. It splits the job into concurrently executing pieces, thereby decreasing the wall-clock run time of the job. Although HPF uses a lot of complicated analysis tools to analyze the user program and can produce a corresponding parallel version without user intervention, the quality of the parallel code produced in such a way is usually very poor. In my study, I found that using

the PGHPF compiler's option -Mautopar to parallelize the MAX3D code automatically, the parallel results are not effective. The compiler is simply not intelligent enough to make a smart decision. Another problem with automatic parallelization is that it can create a large number of temporary arrays. Hence, automatic parallelization has to be combined with hand tuning to produce more efficient code. In my research, only hand parallelization is used to avoid memory waste and increase efficiency.

The HPF model can be seen as a collection of distinguishable, but complementary, programming styles. The models and methods I used in the implementation are data sharing and work sharing. In data sharing, data, such as an array, are spreaded over the memory of all of the available processors so that each processor operates primarily on its own part. In work sharing, the loop iterations are distributed within loops among the system's processors with the goal of executing them in parallel. For instance, you can use the INDEPENDENT directive to divide the iterations of a DO loop among processors, or you can let the compiler divide the work for you by choosing implicit array syntax. One natural and powerful strategy involves distributing an array, data sharing, and the iterations of a DO loop that operate on that array, work sharing, over your available processors. Processors executing a DO loop in parallel help you realize the power of the CRAY T3E and other parallel computing system. In the following sections, the details of data allocation and parallelization schemes are discussed.

2.1 Distributed Data Allocation

In a parallel program, data can be shared or private. Arrays can also be distributed to speedup the code. Arrays are distributed across all processors via the !HPF$ DISTRIBUTE directive in HPF. The DISTRIBUTE directive names the variables that are to be shared data objects and specifies how they are to be distributed across the PEs. If the data is not specified as shared, it is private (the default). In a program that declares a private data item, each processor gets a copy of storage for that item. In many programming models, this data is called replicated. When an array is shared , its elements are distributed across the available PEs.

The DISTRIBUTE directive can specify array distribution within the directive by following the array name with distribution information contained in parentheses. The distribution information may include the keywords BLOCK or CYCLIC, which tell the compiler how to distribute array elements among the available PEs.

The DISTRIBUTE directive allows data to be distributed over processors in a variety of patterns. The ALIGN directive is used to s[ecify that certain data objects are to be mapped in the same way as certain other data objects. Operations between aligned data objects are likely to be more efficient than operations between data objects that are not known to be aligned. The BLOCK distribution distribute a block of consecutive memory locations to a processor. In our implementation, U0, U1, U2, F, and G are all aligned and have the same distribution, as shown

below. Hence, we can parallelize the code across dimension K efficiently when working on these large arrays. The ONTO clause specifies the processor arrangement declared in a PROCESSORS directive. In our case, the arrangement is PROCS .

```
!HPF$ PROCESSORS PROCS(NUMBER_OF_PROCESSORS())

!HPF$ ALIGN U1(I,J,K,L) WITH U0(I,J,K,L)
!HPF$ ALIGN U2(I,J,K,L) WITH U0(I,J,K,L)
!HPF$ ALIGN F(I,J,K,L) WITH U0(I,J,K,L)
!HPF$ ALIGN G(I,J,K,L) WITH U0(I,J,K,L)
!HPF$ DISTRIBUTE U0(*,*,BLOCK,*) ONTO PROCS
```

Similarly, since we need to parallelize HZETA on dimension J, the H array is distributed on dimension J as follows:

```
!HPF$ DISTRIBUTE H(*,BLOCK,*,*) ONTO PROCS
```

Since different phases require different data distributions, it would be nice to dynamically distribute an array during execution. HPF provides such mechanism. They are DYNAMIC and REDISTRIBUTE directives. In our research, we also tried to use these two directives to distribute arrays dynamically. Since redistribution of large arrays such as U0 and H take a lot of time, the overhead incurred is lager than the saving in time. Hence, we did not use this redistribution scheme in our final implementation. Another distribution method is CYCLIC. This distribution distributes an array of elements cyclically to different processors. We found this method was not efficient for the MAX3D code.

2.2 Parallelization

The major part of the parallelization process for the MAX3D code is loop parallelization since the codes contain several big loops. The model of parallelism used focuses on the Fortran DO loop. The compiler executes different iterations of the DO loop in parallel on multiple processors.

The essential compiler directive for multiprocessing in HPF is !HPF$ INDEPENDENT. This directive directs the compiler to generate special code to run iterations of a DO loop in parallel. The loop is also called shared loop. In a shared loop, the iterations are divided among the available tasks. A shared loop can be specified by the INDEPENDENT directive. A private loop, by contrast, is executed only by the task that invokes it; no work is shared among tasks. Private loops are not preceded by an INDEPENDENT directive. For example, the following

directive specifies that loop K in subroutine FXI should be parallelized and I, IM, etc, are private variables.

```
!HPF$ INDEPENDENT, NEW(I, IM, IP, J, SSXI, RSSXI, UI1, UI2, UI3, RX, RY,&
!HPF$&RZ,                                                               &
!HPF$&UJX, UJY, UJZ, SUJ, RSUJ, UJ1, UJ2, UJ3, UK1, UK2, UK3,          &
!HPF$&UP1, UP2, UP3, UP4, UP5, UP6,                                    &
!HPF$&UC2, UC3, UC5, UC6, FP2, FP3, FP5, FP6, FP, FM2, FM3, FM5,       &
!HPF$&FM6, FM)
      DO 1 K=1,KLM
      DO 1 J=1,JLM
      DO 2 I=1,ILM
```

Shared loops specify the behavior of all tasks collectively, but they define the behavior of individual tasks only implicitly. Shared loops do not guarantee the order in which iterations will be executed. The lack of a defined order lets the system execute iterations concurrently.

Inside a shared loop, each task executes its assigned loop iterations as if each task were its own serial region. The assignment of loop iterations is accomplished by aligning the iterations according to the distribution of the array named in the ON clause of the INDEPENDENT directive.

There is an implicit barrier synchronization at the end of a shared loop, at which all tasks wait until the last one has completed. The shared loop ends after the DO loop that immediately follows the INDEPENDENT directive. After the shared loop has finished executing, the tasks continue to execute in parallel.

For multiprocessing to work correctly, the iterations of the loop must not depend on each other; each iteration must stand alone and produce the same answer regardless of when any other iteration of the loop is executed. Not all DO loops have this property, and loops without it cannot be correctly executed in parallel. However, many of the loops encountered in practice fit this model. Further, many loops that cannot be run in parallel in their original form can be rewritten to run wholly or partially.

Many loops that have data dependencies can be rewritten so that some or all of the loop can be run in parallel. The essential idea is to locate the statement(s) in the loop that cannot be made parallel and try to find another way to express it that does not depend on any other iteration of the loop. Through loop merge and loop switch, we eliminate a lot of data dependencies in the code. Some data dependencies can not be eliminated due to the nature of the computations involved in the code. Our strategy is to parallelize the loops which do not have any data dependencies. For example, loop K in subroutine HZETA has data dependen-

cies. However, loop J does not. Hence, we parallelized loop J instead of loop K in subroutine HZETA. Similarly, we parallelized loop K in subroutines FXI and GETA. This is also reflected in their data distribution: array F, G, and H are distributed on different dimensions as shown previously.

3 Performance Results

During this research, I used the CRAY T3E computer system at the Ohio Supercomputer Center, which is a powerful scalable parallel system with 128 processing elements. Its Peak performance can reach 76.8 GFLOPS. Each processor is a DECchip 64-bit super-scalar RISC processor. It has four-way instruction issue with two floating-point operations per clock. Each processor has on-chip 8 Kbyte direct-mapped L1 instruction cache and on-chip 8 Kbyte direct-mapped L1 data cache. It also has on-chip 96 Kbyte three-way-set-associative L2 unified cache. Each processor has a local memory of 16 Mwords (or 128 Mbytes). The clock speed is of the processor is 300 MHz and the peak performance of the processor is 600 MFLOPS. Although the speed is quite fast, compared with other parallel systems such as the SGI Origin 2000, the cache and local memory sizes of the T3E are much smaller (Each processor in the SGI Origin 2000 contains a 4 Mbytes secondary cache). This limits the power of the T3E for programs using a lot of memory space such as the MAX3D code. Actually, the effect of smaller cache size on the T3E is not only on parallel programs, but also on sequential programs. Although the processor speed in the T3E is higher than those in the SGI Origin 2000 or the IBM SP2, the system performce of the T3E is no better than the SGI Origin 2000 as shown below in our research.

Several experiments are carried out to tune our parallel code and to adjust our strategy as to how to distribute the various arrays and which loops and subroutines to parallelize. In this section, some experimental results are reported for the final parallel code produced. All the times are in seconds in the following presentation.

Table 1 lists the the timing information for the major subroutines in the code. The corresponding times are also presented in Figure 1. Since these subroutines are repeated many times, the time spent there contributes heavily to the total time of the code. As shown in Figure 1, all the subroutines are scalable up to 98 processors. When using 128 processors, no further reduction in time is achieved. It is also clear that the subroutine SUM spends the most time compared to other subroutines. Due to memory size, the program cannot be executed on the T3E with fewer than 4 processors.

Table 2 shows the execution times the code after loop transformations and hand parallelization. Because of the limited memory space available on each processor in the T3E system, the parallel program could be run on the T3E with one or two processors. Hence, Table 2 only shows the execution times using at least four processors. Due to the requirement of the HPF

No. of Processors	FXI	GETA	HZETA	SUM
4	0.59236	0.89535	0.8809	2.8894
8	0.31400	0.45646	0.44389	1.4939
16	0.16963	0.25290	0.22604	0.83692
32	0.099733	0.15042	0.11862	0.50716
48	0.074417	0.10890	0.11501	0.37680
64	0.053583	0.076501	0.064894	0.25680
90	0.053734	0.07640	0.064764	0.25511
98	0.02948	0.04119	0.06406	0.1443
128	0.02948	0.04108	0.06384	0.1430

Table 1: Execution Times in the Major Subroutines.
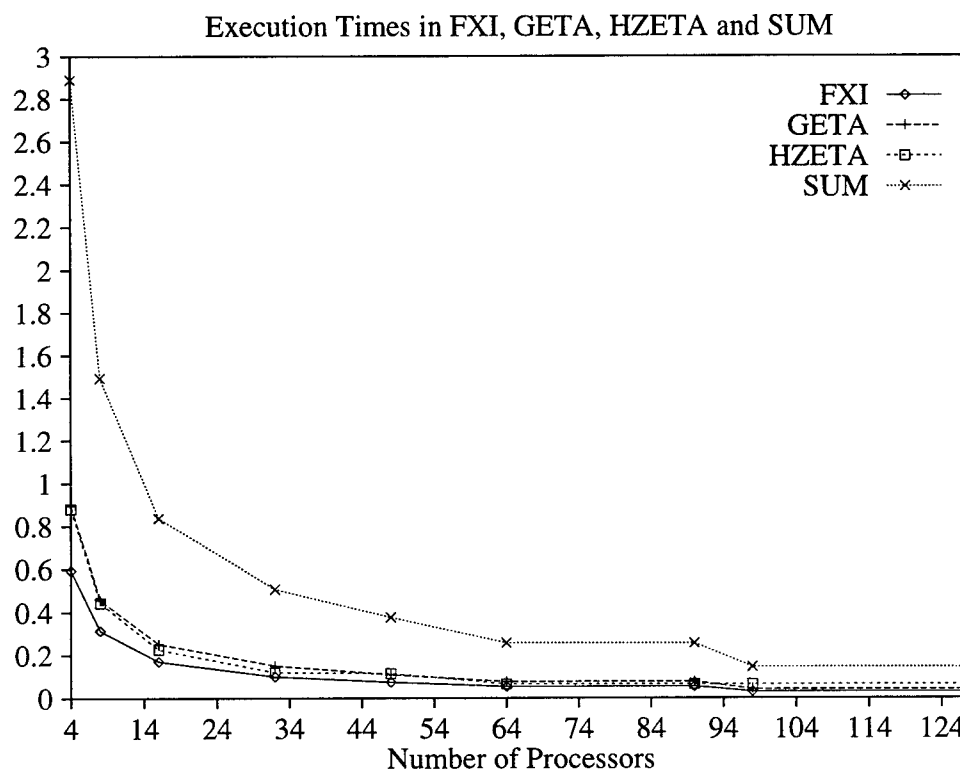
Figure 1: Execution Times of the Major Subroutines.

No. of Processors	Execution Time
4	16331
8	8713
16	4853
32	2919
64	1659
71	1680
90	1688
98	1127
128	1148

Table 2: Total Parallel Execution Times for Problem Size $(73 \times 61 \times 97)$.

compiler I used, many huge arrays have to be initialized before their usage, a lot of overhead is introduced. Hence, the execution time on the T3E with four processors is larger than the execution time on the Origin 2000 with four processors. However, the HPF code on the T3E is more scalable than the PFA code on the Origin 2000. From Figure 2, we can also see that the execution time reaches its minimum when 98 processors are used. Clearly, the parallel code not only reduces the total execution time drastically, but also is scalable up to 98 processors on the T3E system.

We can also see that the time is not reduced when we increase the number of processors from 64 to 71 or 90. This is natural since we parallelize the code on dimension K which has a size of 97. 71 or 90 processors are not enough to partition the array on dimension K. Due to array padding, the actual size on dimension K is 98 and hence has a perfect partition over 98 processors. Increasing the processor number is not effective unless we can use 196 processors. We predict that the execution time can be further reduced by using 196 processors. Unfortunately, the T3E system has only 128 processors.

Because the timing information is not available on the T3E when using only one processor, we cannot calculate the speedup exactly. In order to estimate the speedup, we assume that the program is executed optimally on the T3E with four processors. This assumption is quite reasonable since experiments indicate that the MAX3D code is run efficiently with a small number of processors. With this assumption, we can estimate the execution time on the T3E with one processor via multiplying the execution time on the T3E with four processors by four. Table 3 shows the speedups of the HPF code after loop transformations and hand parallelization using the above assumption. The corresponding data is also shown in From Figure 3. It is clear that the code is still quite scalable when the number of processors used is 98.
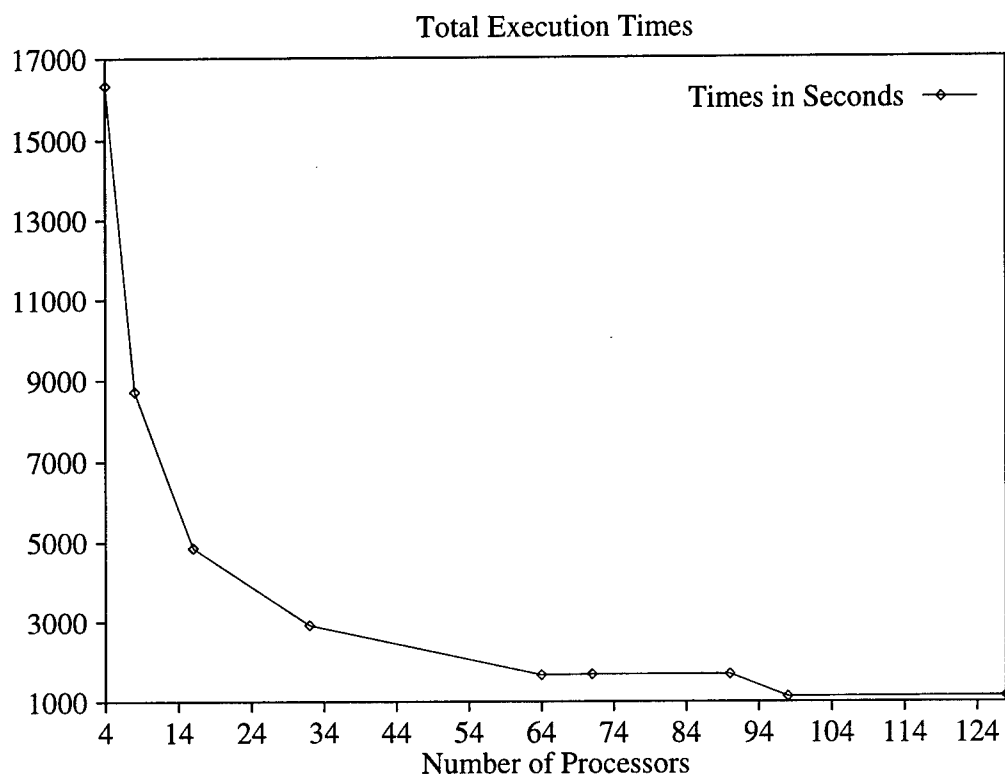
Figure 2: Execution Times of the HPF Code.

No. of Processors	Speedup
1	1
2	2
4	4
8	7.49
16	13.46
32	22.38
64	39.38
71	38.88
90	38.70
98	57.96
128	56.90

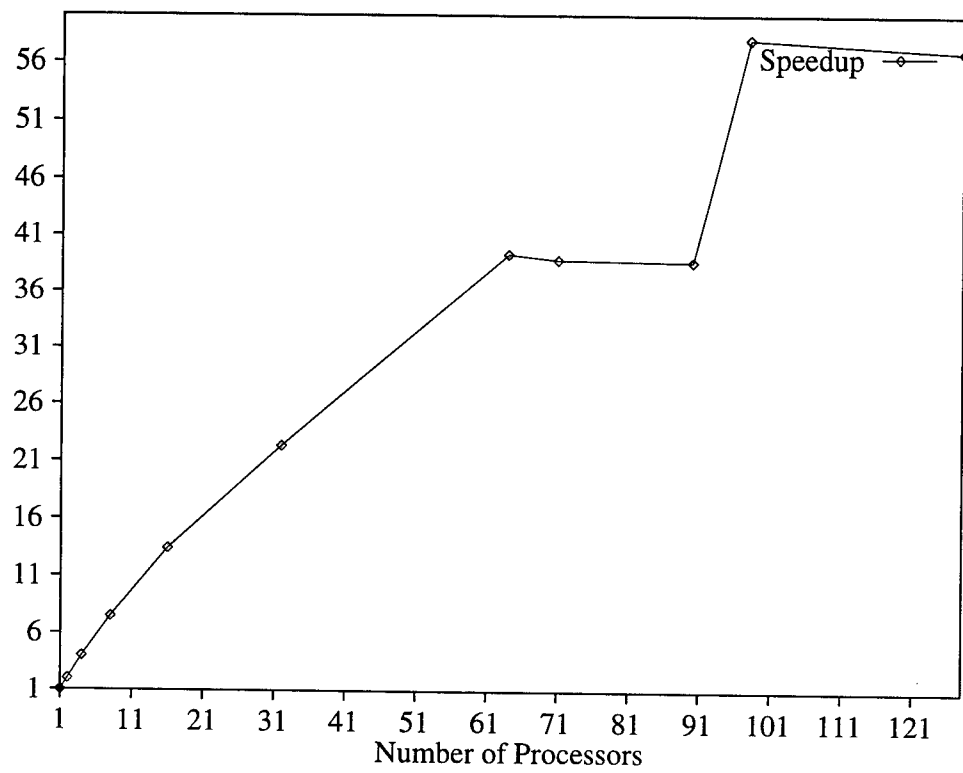Table 3: Speedups for Problem Size $(73 \times 61 \times 97)$

Figure 3: Speedups for Problem Size $(73 \times 61 \times 97)$

The experiments conducted in this research are limited to a fixed problem size ($73 \times 61 \times 97$). When the problem size changes, the scalability and efficiency may also change. Theoretically, when the problem size is bigger, the ratio of computation over communication is also getting bigger. This implies that the communication overhead becomes relatively smaller. It is our conjecture that the parallel code will be more scalable and efficient when computing larger problems. However, more experiments are needed to confirm our conjecture.

4 Conclusions

As we all know, the message passing paradigm has several disadvantages: the cost of producing a message passing code may be between 5 and 10 times that of its serial counterpart, the length of the code grows significantly, and it is much less readable and less maintainable than the sequential version. Most importantly, the code produced using the message passing paradigm usually uses much more memory than the corresponding code produced using high level parallel languages since a lot of buffer space is needed in the message passing paradigm. For these reasons, it is widely agreed that a higher level programming paradigm is essential if parallel systems are to be widely adopted. The results obtained in my summer research show that the performance of the parallel PFA code is even better than the MPI counterpart besides the advantages stated above. However, the current version is implemented in Power Fortran, and can only be run on SGI Power Challenger and Origin 2000. The current PFA version also has the problem of not scalable when the the number of processors is larger than 16. On the other hand, HPF is more powerful than Power Fortran and is available in almost every major parallel computer systems.

HPF has reached a critical stage in its history. Having struggled while the compiler technology evolved into a usable state, the parallel computing community has now found it possible to write portable, high-performance implementations for selected applications in HPF. This research indicates that large scale data-parallel applications such as CEM simulations can use HPF to achieve reasonable performance. However, there are a number of pitfalls ahead. Notable among these are the general drift away from Fortran for advanced applications, remaining difficulties with HPF optimization technology, inadequacy of support for irregular problems in current implementations, the continuing need for advanced mathematical library support, and the emergence of competing paradigms such as MPI and PVM.

This research addresses the portability and scalability problems of the MAX3D code through using HPF. Our results achieved during the research have demonstrated that the sequential MAX3D code can be parallelized much more quickly using HPF than using MPI, and the doce produced this way can still be executed efficiently. Further, the code is still quite scalable. Although the HPF code is running more slowly than the corresponding MPI code and the PFA

code on the SGI Origin 2000, the problem will be alleviated in the future or running on other machines. The problem is partially due to the HPF compiler called PGHPF I used during this research. The PGHPF High Performance Fortran compiler supports all HPF 1.1 extensions to Fortran 90 and a growing number of HPF 2.0 features [19]. The PGHPF compiler version I used is 2.3 and it requires that all arrays be initialized before their usage. It also has some other limitations and bugs which will be eliminated in the future. On the other hand, MPI or PFA do not have the initialization requirement. This could save a lot of time when a large number of big arrays are used in the code. Another possible reason is because processors on the T3E has snall memory and cache space. Hence, the HPF code on one processor is slower than its counterparts. Nevertheless, the scalability of the HPF code is quite good and is better than that of the PFA code and close to that of the MPI code. As newer version comes out and HPF compilers become more mature, we expect that the performance of the HPF MAX3D code will be improved substantially in the near future with little change in the code.

5 Acknowledgements

References

[1] Joseph S. Shang and Datta Gaitonde and K. Wurtzler. Scattering Simulations of Computational Electromagnetics. 27th AIAA Plasmadynamics and Lasers Conference, AIAA-96-2337, AIAA, 1996.

[2] Joseph S. Shang and Datta Gaitonde. High-Order Finite-Volume Schemes in Wave Propagation Phenomena. 27th AIAA Plasmadynamics and Lasers Conference, AIAA 96-2335, AIAA, 1996.

[3] Joseph S. Shang and Datta Gaitonde. On High Resolution Schemes for Time-Dependent Maxwell Equations. 34th Aerospace Sciences Meeting and Exhibit, AIAA 96-0832, AIAA, 1996.

[4] Joseph S. Shang and Robert M. Fithen. A Comparative Study of Characteristic-Based Algorithms for the Maxwell Equations Journal of Computational Physics, 125: 378 - 394, 1996.

[5] Joseph S. Shang and S.J. Scherr. Time-Domain Electromagnetic Scattering Simulations on Multicomputers. 26th AIAA Plasmadynamics and Lasers Conference , AIAA 95-1966, AIAA, 1995.

[6] Joseph S. Shang and D.A. Calahan and B Vikstrom. Performance of a Finite Volume CEM Code on Multicomputers. Computing Systems in Engineering, 6(3): 241 - 250, 1995.

[7] Joseph S. Shang and Datta Gaitonde. Scattered Electromagnetic Field of a Re-Entry Vehicle. Journal of Spacecraft and Rockets, 32(2) : 294 - 301, 1995.

[8] Joseph S. Shang and Datta Gaitonde. Characteristic-Based, Time-Dependent Maxwell Equation Solvers on a General Curvilinear Frame. AIAA Journal, 33(3): 491 - 498, 1995.

[9] Joseph S. Shang. A Fractional-Step Method for Solving 3-D Time-Domain Maxwell Equations. 31th Aerospace Sciences Meeting and Exhibit, AIAA 93-0461, AIAA, 1993.

[10] Joseph S. Shang. Characteristics Based Methods for the Time-Domain Maxwell Equations. 29th Aerospace Sciences Meeting, AIAA 91-0606, AIAA, 1991.

[11] J.S. Shang, M. Wagner, Y. Pan and D.C. Blake, "Strategies for time-domain CEM computations on multicomputers," 36th Aerospace Sciences Meeting, Reno, NV, January 12-15, 1998, AIAA Reprint 98-0979.

[12] Michael Wolfe. High Performance Compilers for Parallel Computing, Addison-Wesley Publishing Company, 1996.

[13] James Laudon and Daniel Lenoski. The SGI Origin: A ccNUMA Highly Scalable Server, the 1997 International Symposium on Computer Architectures, Denver, Co, pp. 241-251.

[14] Chris Hogue. MIPSpro (TM) Power Fortran 77 Programmer's Guide, Silicon Graphics, Inc. 1996.

[15] Marc Snir, et al. MPI : the complete reference, MIT Press, Cambridge, Mass., 1996.

[16] W. Gropp, E. Lusk, A. Skjellum. Using MPI : portable parallel programming with the message- passing interface, MIT Press, Cambridge, Mass., 1994.

[17] C. H. Koelbel. The High performance Fortran handbook, MIT Press, Cambridge, Mass., 1994.

[18] K. Hwang and Z. Xu. Scalable Parallel Computing, McGraw-Hill Companies, Inc., 1998.

[19] Douglas Miles, Vincent Schuster, and Mark Young. The PGHPF High Performance Fortran Compiler: Status and Future Directions, The 2nd Annual HPF User Group meeting, Porto, Portugal, 25 - 26 June 1998.

A Hybrid Variational-Asymptotic Method for the Analysis of Micro-Mechanical Damage in Composites

Kishore Pochiraju
Principal Investigator
Department of Mechanical Engineering
Stevens Institute of Technology
Castle Point on Hudson
Hoboken, NJ 07030

Laboratory Focal Point
Dr. Nicholas J. Pagano Jr.
AFRL/MLBC
Wright Patterson Air-Force Base
WPAFB, Dayton, OH

January, 1999

Table of Contents

1. SUMMARY OF EFFORT AND OUTCOMES

This SREP extension focuses on the development of a refined methodology to determine stress distributions in the presence of local material and geometric discontinuities, such as cracks and corners in composite materials. The belief that micro-mechanical damage initiates or propagates in such regions motivates this research. The outcomes of this research extension are as follows:

- Development and Refinement of the ADM (Variational)-Asymptotic combination methodology (Section 2).

- Determination of the matching radius for the ADM and asymptotic solutions and the generalized stress intensity factors (Section 3).

- Assessment of the effectiveness of the developed solution methodology through case studies (Section 4).

- Enhancement of the ADM software to include asymptotic analysis and the ADM-Asymptotic solution matching technique. CDROM of enhanced software provided in copy sent to the lab focal point, (Dr. N. J. Pagano, AFRL/MLBC).

- Continued collaboration with AFRL/MLBC researchers on finite element analysis and singular stress fields analysis outside of this project scope --- Fiber break simulation and cruciform simulation.

The project effort and outcomes correspond with those anticipated in the proposal. These results were presented to the laboratory focal point and a team of researchers on two occasions. A mid project meeting was held at AFRL, Wright Patterson Air Force Base in June 1998. At this meeting the results of the project were reviewed and further work was planned. (Attendees: Drs. Pagano and Tandon of MLBC/AFRL, K. Pochiraju and J. Charvet of Stevens). A second meeting was held during the ASME winter conference at Anaheim, CA with Drs. Pagano, Tandon, and Pochiraju in attendance. The project outcomes were reviewed by the laboratory focal point and the contents of this report were presented to the AFRL researchers. Subsequently, the software developed under this project was delivered.

The funds allocated in this project were used for supporting one graduate student for two semesters, for institutional overhead, and for travel. The travel included ASME

winter conference and to the AFRL, Wright Patterson Air Force Base, for the status meeting.

The results of this project are documented in several technical conferences. Journal articles based on this work are currently prepared for submission. The Appendices A and B contain the presentation material and the manuscript being prepared for journal submission.

2. THE HYBRID-VARIATIONAL ASYMPTOTIC METHOD

In this section a technique for embedding the asymptotic solutions within selected regions of the variational model [1-5] is described. This technique modifies the variational solution while ensuring that the global equilibrium, boundary conditions and the constitutive relationships are satisfied. Stress continuity between the asymptotic and variational solution at one selected point inside the domain is matched to determine the unknown coefficients in both ADM and variational models. Appendix – A provides a detailed report with the background on the development of the model and the formulations.

Consider an axisymmetric model of fiber pullout specimens as shown in Figure 1. In obtaining the refined solution, a region of radius ρ^*, is identified around the singular point (corner-O in Figure 2). In this region the asymptotic solution [6,7] determined by solving the local homogenous problem is taken to dominate and the variational solution in this region is ignored. The local asymptotic solution, however, can only be determined with unknown scale coefficient(s). This scale factor needs to be determined using a global solution.

Since a small region around the singular point is excluded from the variational solution, sufficient boundary conditions do not exist for solving the global boundary value problems. Additional equations are required to obtain the undetermined coefficients in the variational formulation. In this methodology, these additional equations are obtained by matching the asymptotic and variational solutions at one point near the point of singularity.
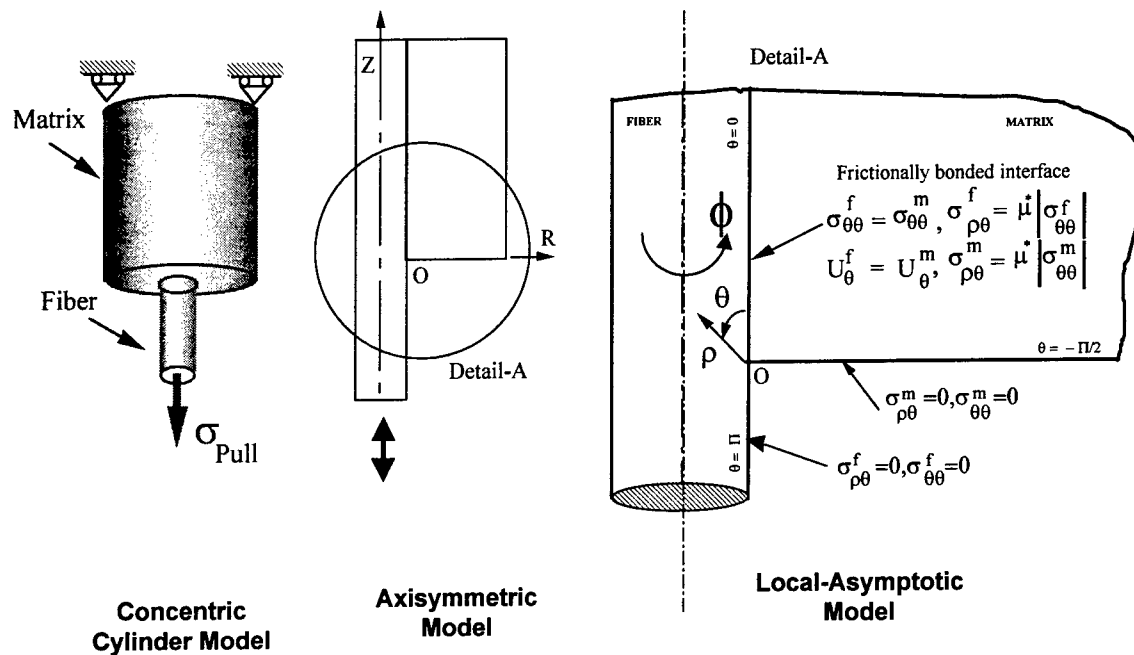
Figure 1: Concentric Cylinder Model - Global axisymmetric and local asymptotic problems

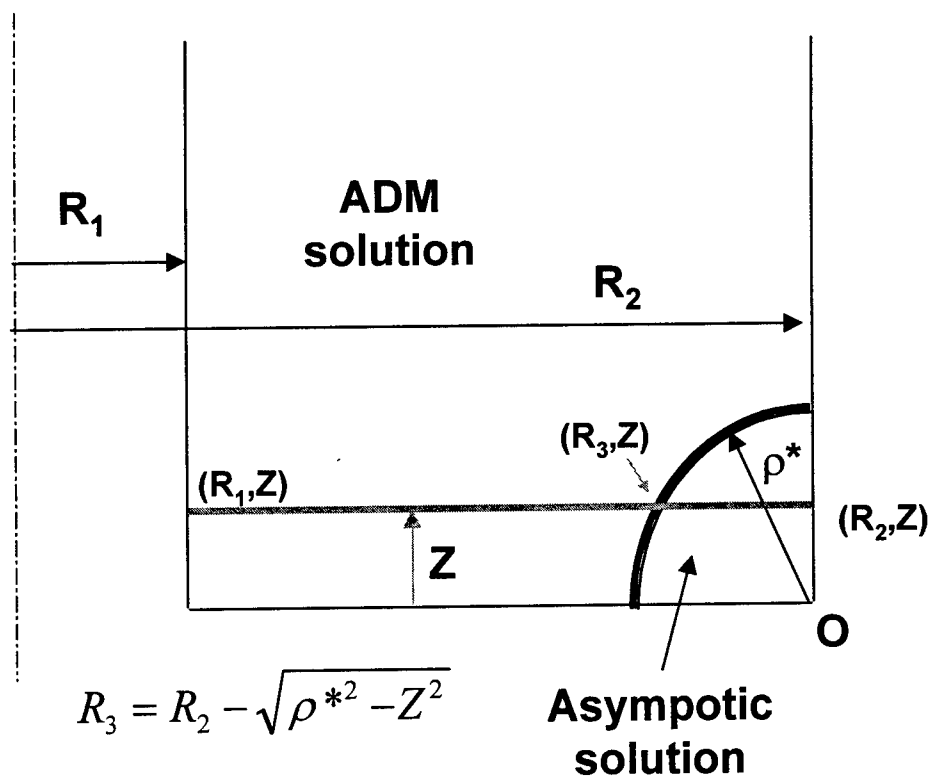$$R_3 = R_2 - \sqrt{\rho^{*2} - Z^2}$$

Figure 2: Regions of applicability of asymptotic and ADM solutions

This point on the boundary between the variational and asymptotic solution is identified with radial and axial coordinates, ρ^*, θ^*, respectively. At this point, conditions are imposed such that the normal, shear and radial stresses are continuous between the variational solution and the asymptotic analysis. If there are 'M' conditions required for the variational solution (thus determining M undetermined constants), and a 'N' term expansion of the asymptotic solution is considered, then N+M independent matching conditions are required to determine both the local and global solutions. For example in Problem-A, there are three boundary conditions associated with the singular point in the variational formulation. They are the axial stress in the fiber (σ_{ZZ}^{F}), the axial stress in the matrix (σ_{ZZ}^{m}), the shear stress (σ_{RZ}, continuous on fiber and matrix side) vanish. These conditions determine three unknown coefficients in the variational solution. If a one term asymptotic solution is considered, then one undetermined scale constant needs to be obtained by matching. Thus there are four unknown coefficients requiring four matching conditions. Following four conditions can be written at the matching radius, ρ^*, following conditions at $R=R_f$, $Z= Z^*(Z^* \neq 0)$:

$$\sigma_{ZZ} \,(ADM) = \sigma_{ZZ} \,(ASYMP) \text{ in fiber}$$

$$\sigma_{ZZ} \,(ADM) = \sigma_{ZZ} \,(ASYMP) \text{ in matrix}$$

$$\sigma_{RZ} \,(ADM) = \sigma_{RZ} \,(ASYMP)$$

$$\sigma_{RR} \,(ADM) = \sigma_{RR} \,(ASYMP) \qquad (1)$$

First issue addressed is the determination of the size of the local region where the asymptotic solution prevails. In the current parameterization of the problem, this translates to determining the value for the radius, ρ^*. An error-function [7] is employed to obtain the difference in stress fields predicted by asymptotic and variational solutions at several radii. Using such an error-basis, an optimal radius for embedding the asymptotic solution within the variational solution is searched.

The error function used to quantify the extent of angular variation matching at radius, ρ^* is given in eq.(2). The scale factor (K) in the error computation is obtained using a single point matching as given in eq. (3).

$$E = \sqrt{ \frac{ \sum_{p=1}^{N} \left\{ \left[\frac{ \left\{ \frac{\sigma_{\rho\rho}\left(\overline{\rho},\theta_p\right)^{ADM}}{k\overline{\rho}^{\lambda}} - \widetilde{\sigma}_{\rho\rho}\left(\theta_p\right)^{asymp} \right\}^2 + \left\{ \frac{\sigma_{\theta\theta}\left(\overline{\rho},\theta_p\right)^{ADM}}{k\overline{\rho}^{\lambda}} - \widetilde{\sigma}_{\theta\theta}\left(\theta_p\right)^{asymp} \right\}^2 + \left\{ \frac{\sigma_{\rho\theta}\left(\overline{\rho},\theta_p\right)^{ADM}}{k\overline{\rho}^{\lambda}} - \widetilde{\sigma}_{\rho\theta}\left(\theta_p\right)^{asymp} \right\}^2 }{3} \right] \right\}_p }{N} } \qquad (2)$$

The procedures for determining the stress intensity factors (or scale factors for the corner case) is be established. This aspect of the methodology is closely coupled with the first issue of determining the region of dominance for the asymptotic solution. As these scale / stress intensity factors relate the local solution to the global solution of the problem, the determination of the scale factors is best accomplished at the matching point. The selection of a component of the stress field, such as the normal stress component shown in Eq.(3), that most accurately estimates the value of the scale factor is investigated.

$$k = \frac{ \sigma_{\theta\theta}\left(\rho,\theta^{*}\right)^{ADM} }{ \rho^{\lambda}\widetilde{\sigma}_{\theta\theta}\left(\theta^{*}\right) } \qquad (3)$$

In the case of ADM, results indicate that the normal and shear stresses match well closer to the singular point ($\rho^{*} < 1E\text{-}4$ mm) and the radial stress which is discontinuous across the interface matches well away from the singular point ($\rho^{*} > 1E\text{-}2$ mm). As explained in the SFP report submitted in November 1997, this is due to the presence of the boundary condition in the variational solution requiring the radial stress at this boundary to vanish. Thus, the use of the matching technique for the ADM solution is not as straight forward as compared to the FEM solution. In this effort, the error between the ADM and the asymptotic solution and the angular variation matching is analyzed. This analysis identifies that a region of matching where the stress intensity (or scale) factors can be determined accurately. The resuts section shows a numerical example illustrating the procedure.

The issue pertaining to the changes in the interpolation of stresses obtained by the variational solution due to the embedded asymptotic solution is also investigated. Currently the radial interpolation of the stresses within the variational solution is carried out for different components of stress based on the assumed form (the assumed form shown for σ_{zz} and $\sigma_{\phi\phi}$ stress components in Eq. (4)) or the form derived from

equilibrium equations [4]. Figure 2 shows an ADM solution region with radial bounds at R_1 and R_2 and with asymptotic solution embedded around the point of singularity, O. At heights $Z > \rho^*$ of this region, the *unmodified* axial and hoop stresses along the radial line can be interpolated using the functions shown in Eq. (4). The presence of the asymptotic region modifies the state of stress in the neighborhood $(Z < \rho^*)$ of the embedded region. When $Z < \rho^*$, appropriate modifications to Eq. (4) are necessary.

$$\sigma_{zz}(R) = \sigma_{zz}(R_1)\frac{(R_2 - R)}{(R_2 - R_1)} + \sigma_{zz}(R_2)\frac{(R - R_1)}{(R_2 - R_1)} \quad (4)$$

$$\sigma_{\theta\theta}(R) = \sigma_{\theta\theta}(R_1)\frac{(R_2 - R)}{(R_2 - R_1)} + \sigma_{\theta\theta}(R_2)\frac{(R - R_1)}{(R_2 - R_1)}$$

As asymptotic solution replaces the variational solution in the neighborhood of the singularity, the interpolation functions should be appropriately modified. This modification is necessary to determine the angular distribution of stresses around the singular points and to compute the matching error. The modified interpolation scheme, which includes the asymptotic stress fields within appropriate regions, is formulated as given in eq.(5). Using interpolating functions that utilize the asymptotic solution at the boundary of both the regions, (R_3, Z) in Figure 2, are considered. With this assumption, the interpolating functions are written as shown in Eq. (5).

$$\sigma_{zz}(R) = \sigma_{zz}(R_1)\frac{(R_3 - R)}{(R_3 - R_1)} + \sigma_{zz}^{ASYMP}(R_3)\frac{(R - R_3)}{(R_1 - R_3)} \quad (5)$$

$$\sigma_{\theta\theta}(R) = \sigma_{\theta\theta}(R_1)\frac{(R_2 - R_1)}{(R_1 - R_3)} + \sigma_{\theta\theta}^{ASYMP}(R_3)\frac{(R - R_3)}{(R_1 - R_3)}$$

3. SOFTWARE IMPLEMENTATION DETAILS

Two major enhancements to the ADM software are made. First is **the inclusion of Airy-Stress function based asymptotic solution** in the ADM variational environment. Second is **the modification of the variational equilibrium solution** to include the matching conditions shown in eq. (1). Figure 3 shows the two major modifications to the ADM code to implement the functionality described in this document.

Several comparatively minor enhancements were made to the code to enable post-processing of the stresses:

- **Angular variation around points of singularity:** In order to obtain the stress distributions around arbitrary points in an ADM model, a procedure was implemented to obtain stress at any point in the model. This requires identification of the section and layer from which the stresses need to be computed for that location and construction of the solution. This modification allows querying any location in the ADM model to obtain stress components in global axisymmetric and local plane strain model.

- **Logarithmic radial variation:** The capability to print stress fields along arbitrary directions from any point in the model was implemented. The steps for traversal in the radial directions can be both linear and logarithmic. This capability is crucial in examining stress fields at very close distances around the singular points.

During the project period another two substantial enhancements were attempted. First was the determination of **displacement fields** based on the stress fields. This issue remains open. The second is the implementation of **higher order particular solutions**. This enhancement was partially completed. These two enhancements are considered to be beyond the scope of the present project effort and no additional effort was made during the project period. However, the code reflects the partial advances made in these two directions.
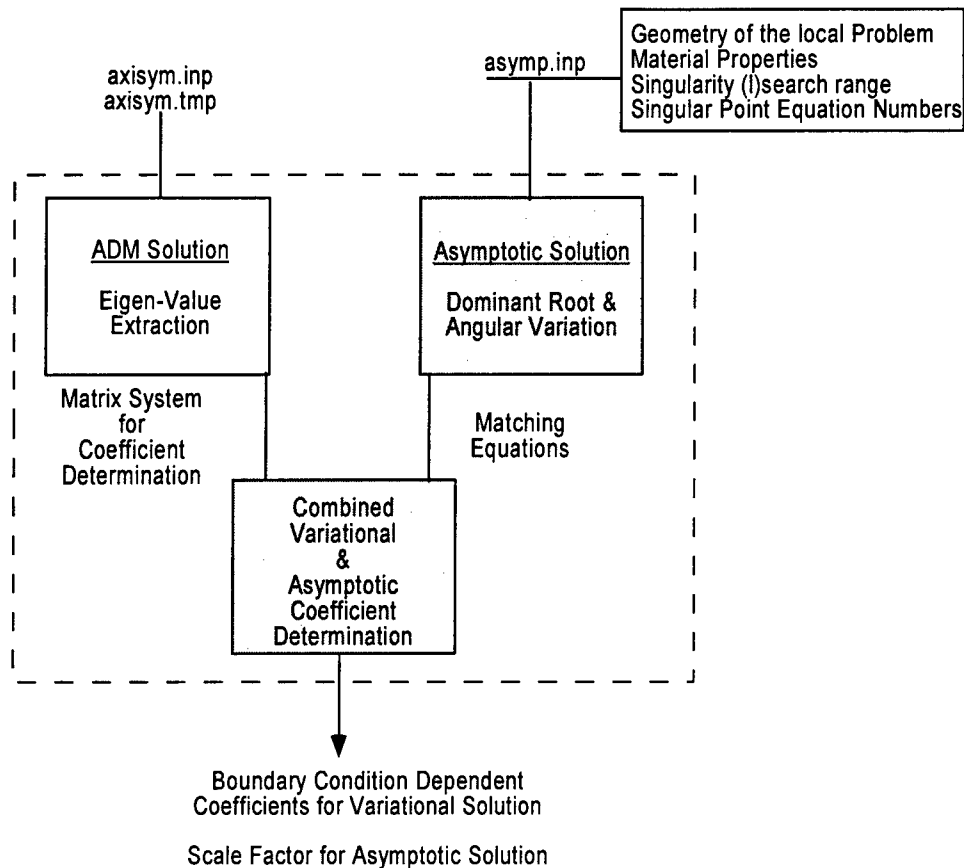
Figure 3: Schematic of additions and modifications made to the ADM Software

(Grey boxes indicate added functionality)

4. RESULTS AND VERIFICATION

Using example problem-A, the original and modified solutions are compared. An arbitrary matching distance of $\rho^* = 1\text{E-}04$ mm is chosen for this example. This problem has a fixed mixity of normal and shear stresses and hence has one undetermined scaling constant for the local asymptotic solution. Three equations are replaced with the matching equations given in Eq. (1). The fourth relation in Eq. (1) is written as an additional equation into the variational equilibrium system. This combined system of global boundary conditions and matching conditions are solved simultaneously to determine the unknown coefficients in the ADM solution and the scale factor of the local solution.

Figure 4-Figure 7 show the axial variation of the original and modified solution plotted along the fiber-matrix interface for several stress components. It can be seen that the solution away from the local singular point is unaffected and the solution in the singular region corrects towards correct asymptotic behavior. This plot also shows that the matching distance from the variational solution perspective should be larger as the influence of the boundary condition (as seen as a local maximum near the singular point in the axial and shear stresses) seem to be of the order of $\rho = 0.1$ mm.
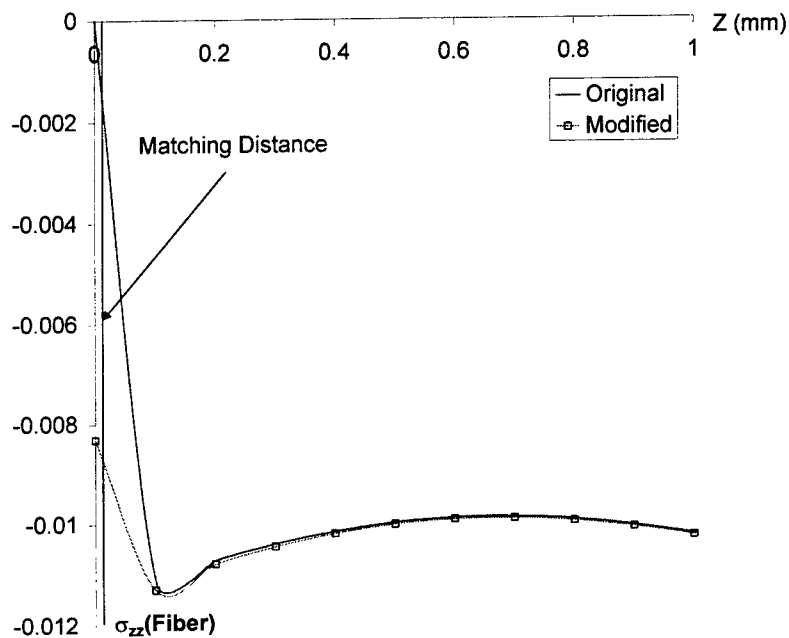

Figure 4: Original and modified solutions for axial stress (Fiber-Side) along the interface
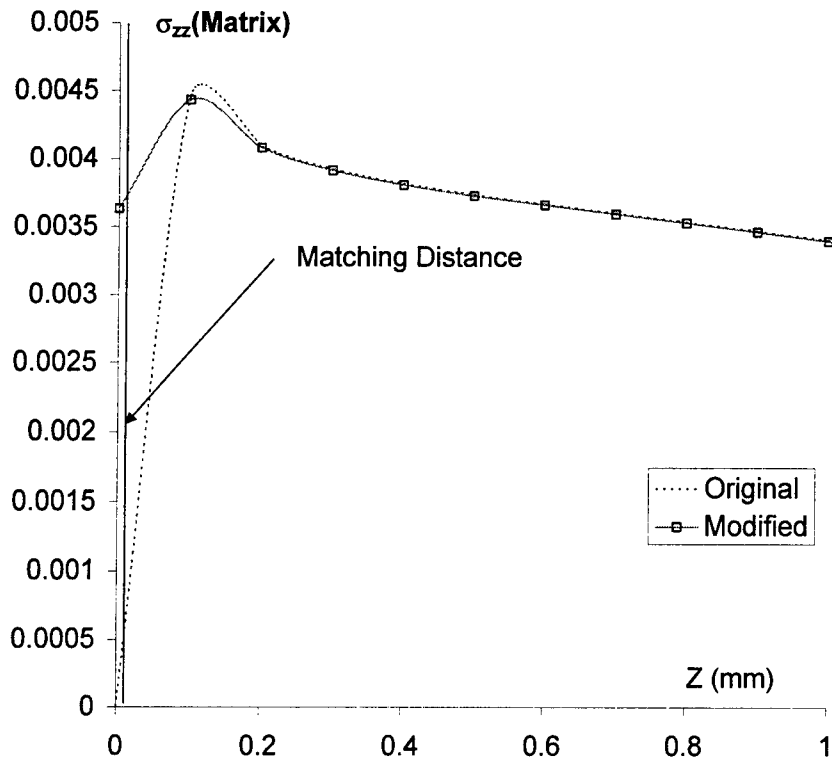
Figure 5: Original and modified solutions for axial stress (Matrix-Side) along the interface
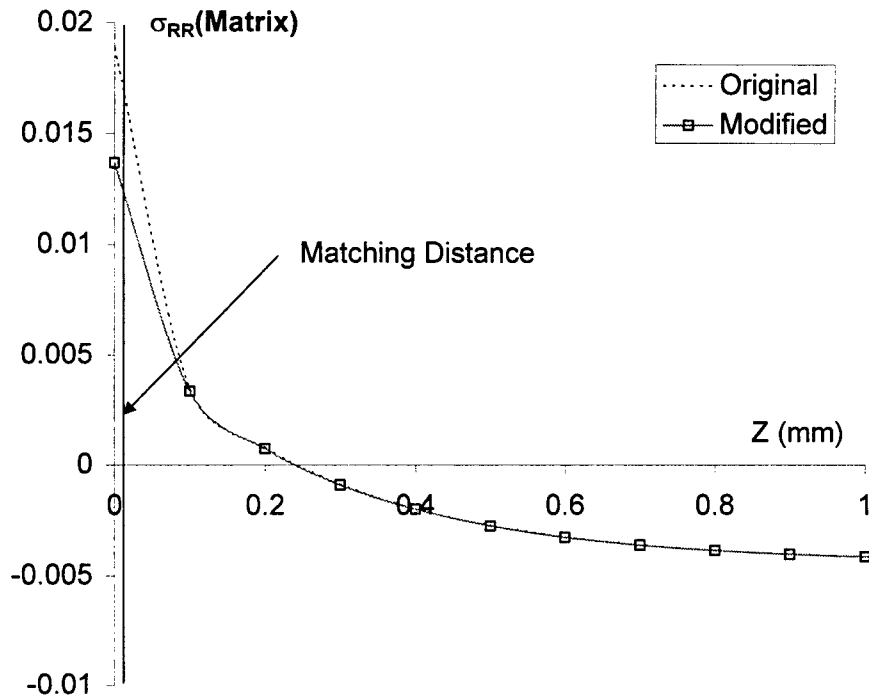

Figure 6: Original and modified solutions for radial stress along the interface
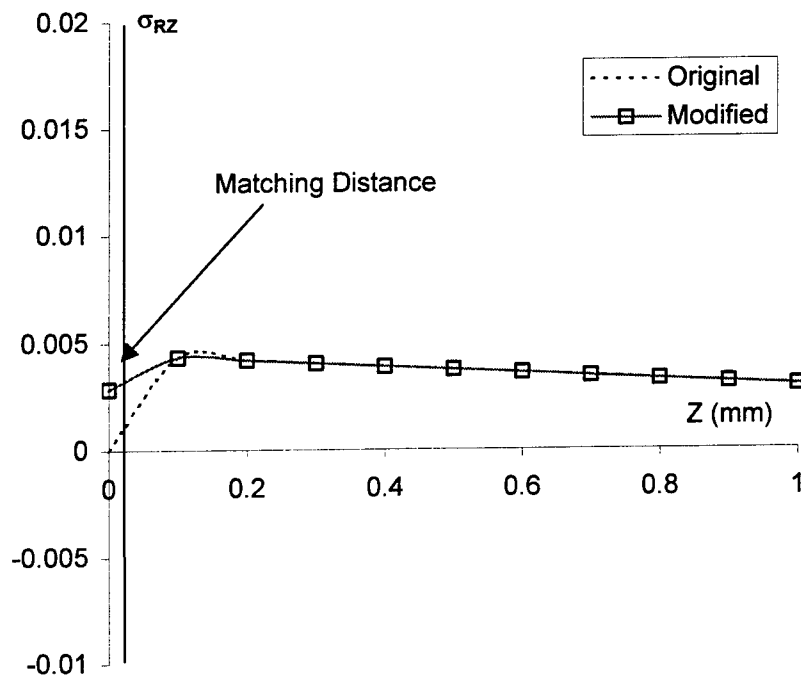
Figure 7: Original and modified solutions for shear stress along the interface

Figures 4 – 7 also show that the modified solution and the original ADM solutions are identical away from the point of the singularity. Thus the modification made only disturbs the solution locally.

The angular variation of the original solution was then compared with that obtained form the asymptotic method. Figures 8-11 show the angular variation matching at 0.0001, 0.001, 0.002, and 0.02 mm away from the singular point. The variational solution is seen to capture two of the three stress components within the meridian plane, namely the opening stress component (σ_{RR}) and the shear stress component (σ_{RZ}) close to the singular point. The axial stress component (σ_{ZZ}), however, can only be matched away from the corner or singular points due to the influence of the boundary condition coincident with the singular point. This behavior is further investigated using the error function.
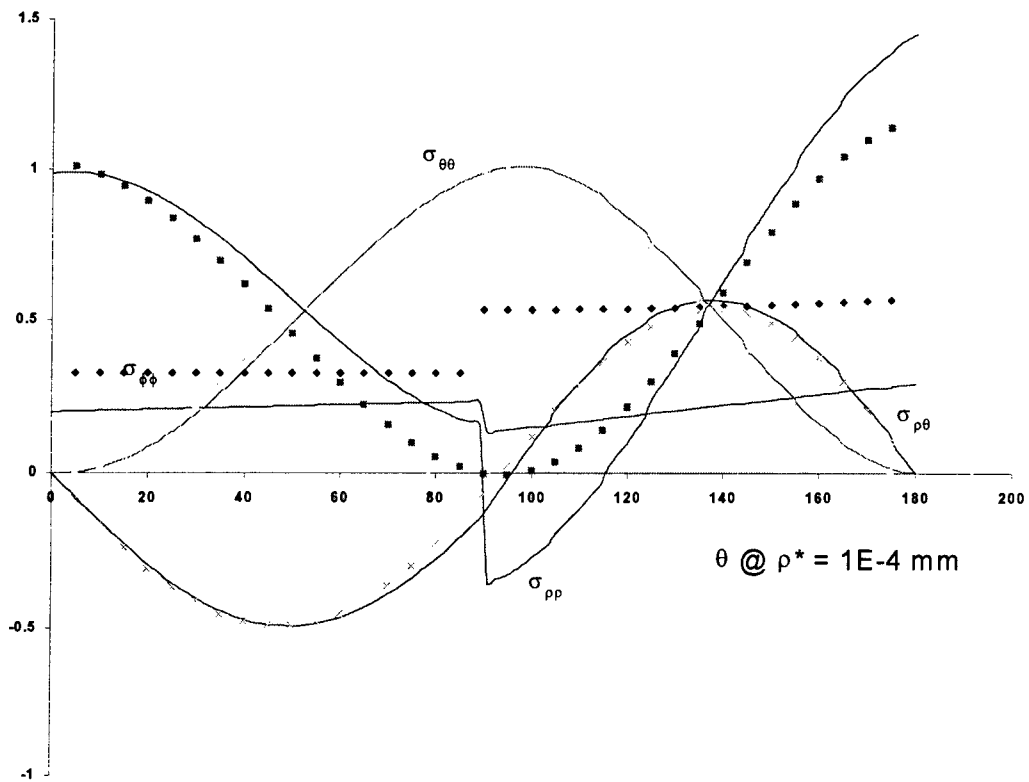
Figure 8: Angular variation matching with ADM solution at ρ*=1E-4 mm
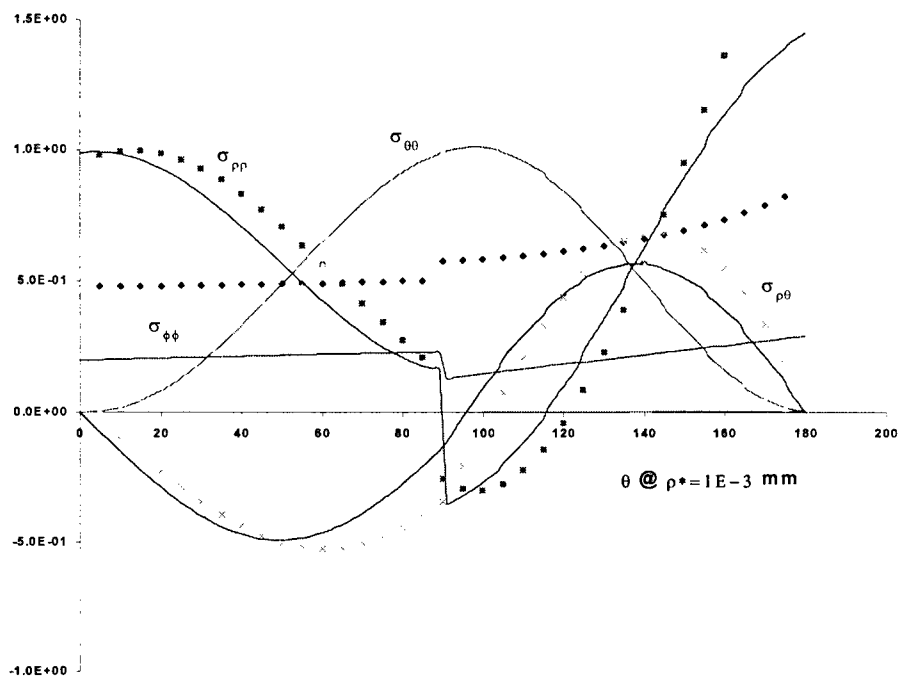

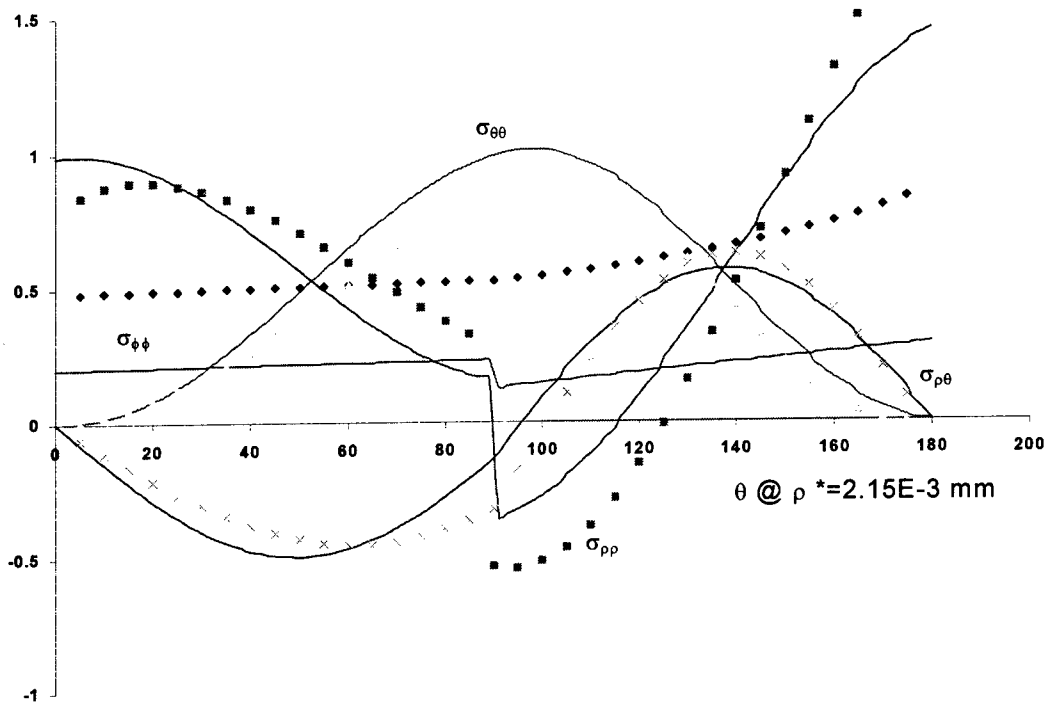**Figure 9: Angular variation matching with ADM solution at ρ*=1E-3 mm**

Figure 10 : Angular variation matching with ADM solution at ρ*=2E-3 mm

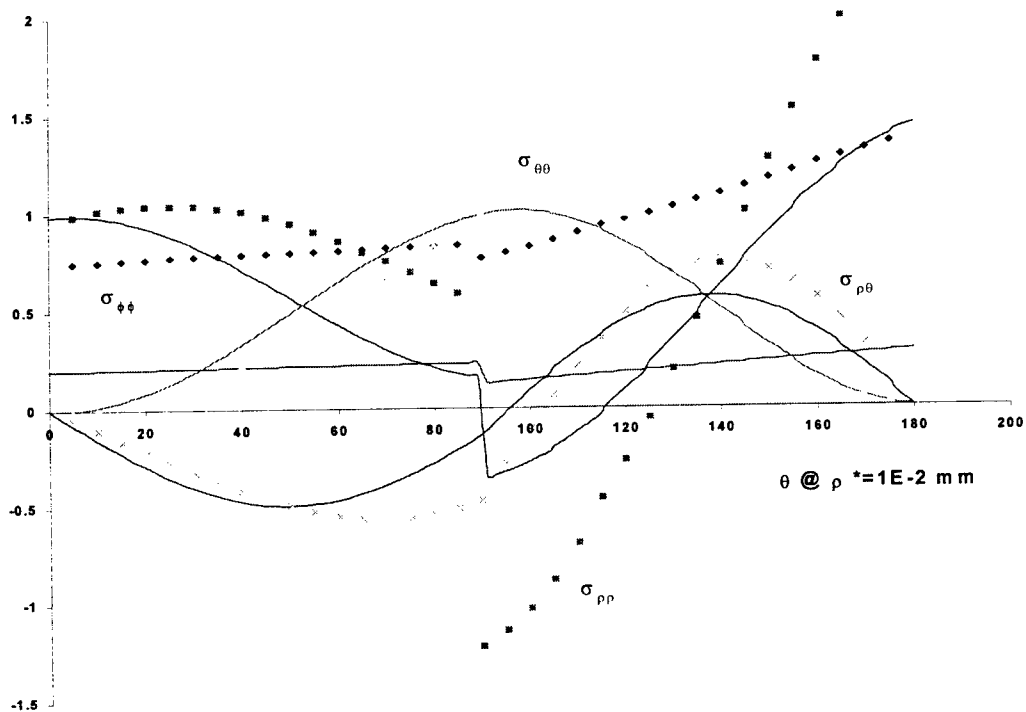**Figure 11: Angular variation matching with ADM solution at ρ*=1E-2 mm**

The error as defined in eq.(2) was computed for the variational solution with the matching distance. Figure 12 shows that the error in the angular variation between the asymptotic and variational solution increases of the matching radius is increases beyond 4.5 μm or reduces below that value. This distance is apparently optimal where the first term in the asymptotic solution is dominant and the radial stress component shows the discontinuity across the interface.

Figure 13 and Figure 14 show the variation of the normal stress and shear stress as determined by the ADM solution near the singular point. The precision of the computational solution is course that no variation of the interface normal stress component($\sigma_{\theta\theta}$) is seen after a distance of 10^{-5} mm. Therefore, a limit on the distance of matching is also imposed by the size of the layering in the ADM methodology. Comparison with the FEM solution shows that the best possible angular variation matching for this case is obtained at a distance of $\rho^* = 8.8E-9$ mm. The angular variations of the asymptotic and FEM solutions match accurately around the point of singularity at this distance. Note that the error in K value grows to 10% at $\rho^* = 3E-7$, and to 44% at $\rho^* = 6.6E-4$. Therefore, more refined variational solutions may be warranted to obtain the scale factors accurately using the matching technique. The modified variational-asymptotic solution determines the scaling factor closer to that of the FEM solution. Figure 13 shows that the hybrid variational asymptotic solution is very sensitive to the matching distance in the region $1E-3 < \rho^* < 1E-2$.

The conclusion that emerges from this investigation is that the optimal matching distance for obtaining the stress intensity and scale factors for variational solutions is farther away from the singular points than those in the FEM solutions. There is a very small region in which this comparison can be effectively made.

Table 1: Variation of Error and Scale Factor K for FEM solution

Matching Radius	K	Error	% Change in K
8.80E-09	0.019731026	0.003984548	0.00
9.65E-09	0.019375432	0.004461569	1.80
1.39E-08	0.019208927	4.51E-03	2.65
3.05E-08	0.018879918	4.98E-03	4.31
6.63E-08	0.018528034	5.48E-03	6.10
3.05E-07	0.017748027	6.81E-03	10.05
1.39E-05	0.015088143	1.22E-02	23.53
0.000663204	0.010926935	2.62E-02	44.62

5. REMARKS

The hybrid variational-asymptotic solution has the potential to be a promising methodology for obtaining stress intensity factors from global variational solutions. However, several challenges emerged during the project period. A conclusion emerges from this work that the region of matching the local and global solutions is further away from the singular point than that in the FEM. *A matching distance where the asymptotic and variational solutions have similar jump at the interface and the hybrid-variational asymptotic method can be used to obtain fairly accurate stress intensity factors.*

Other alternative solutions methods were investigated during the project period. These alternatives and the outcomes of the investigations are briefly summarized in this section.

- Formulation of variational solution with power-type of radial terms: The problem for the formulation arises out of the differences in the local and global solutions. While the global axisymmetric solution is obtained by functions separable in R and Z coordinates, the local solution is separated as functions of ρ and θ. This leads algebraic complexity that seems insurmountable at this point.

- Applicability of Morley's Method: Considerable effort was made to device a method for applying Morley-type of method for embedding radial power-type of terms into the current variational formulation. Morley's method is best applied when the local disturbance can be expressed in terms of a common set of basis functions. It is not obvious to this investigator as to how this can be extended to the situation when a local field is known in terms of one coordinate system and function form and the global field is in a completely different coordinate system and functional form. There may be some difficulties due to the fact that the asymptotic solution used is a plane-strain formulation rather than axisymmetric in nature.

- The displacement continuity, higher order particular solutions, and the more refined layering schemes are refinements that can lead to accurate solutions near the points of singularity.
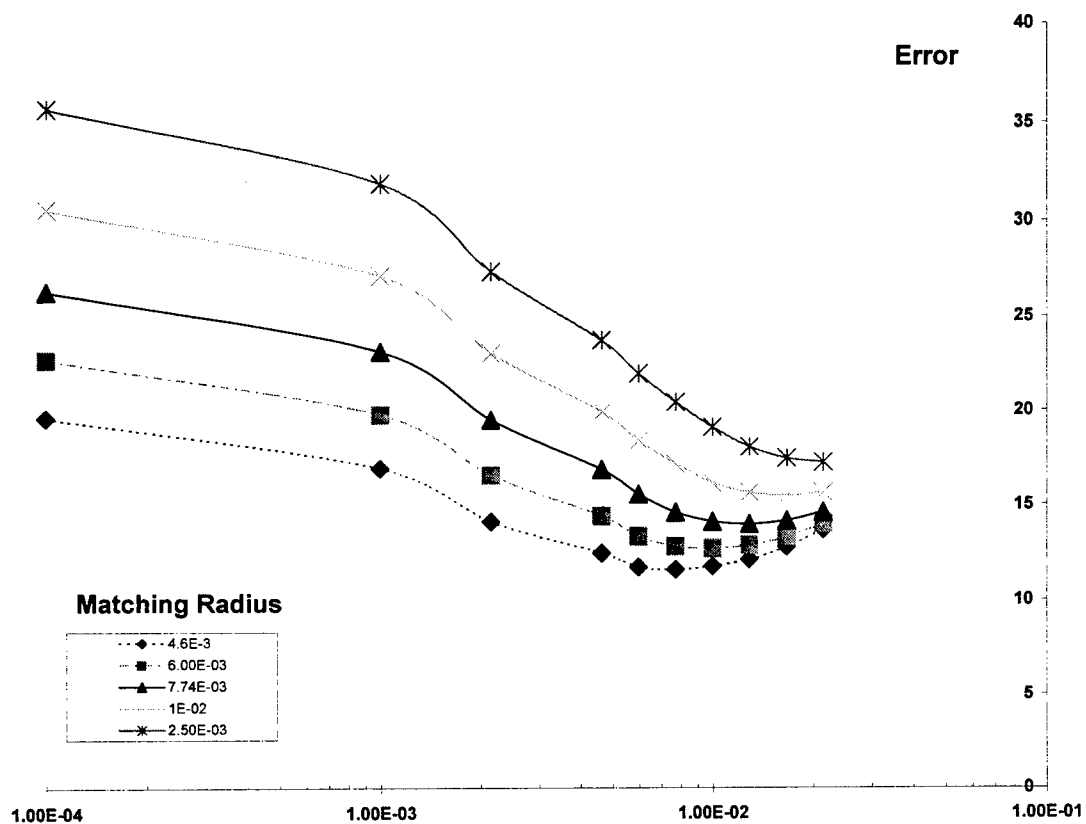
Figure 12: Variation of error in angular variation matching. The range 1E-3 < ρ* < 1E-2 is identified as a region for matching for this problem.
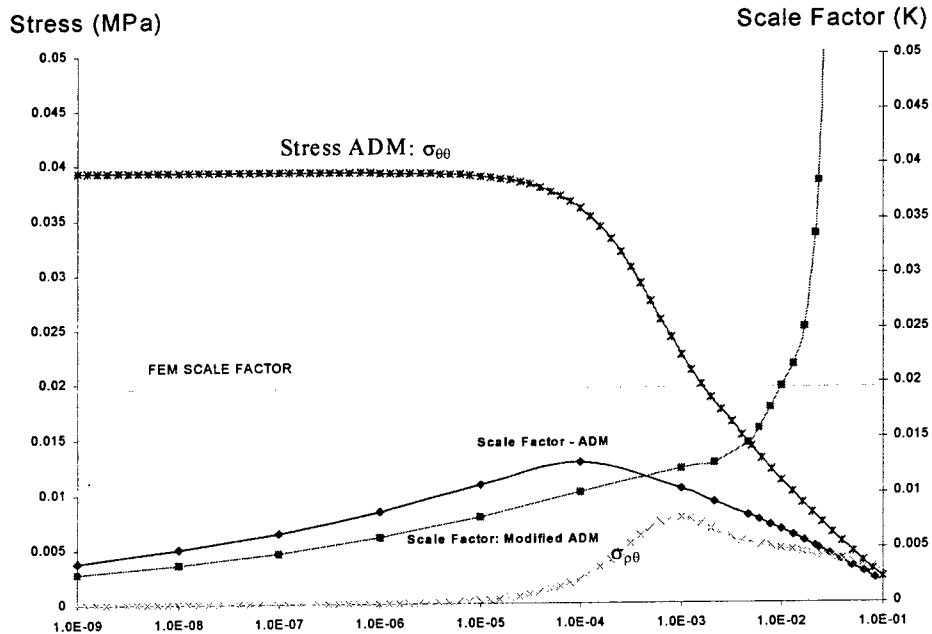
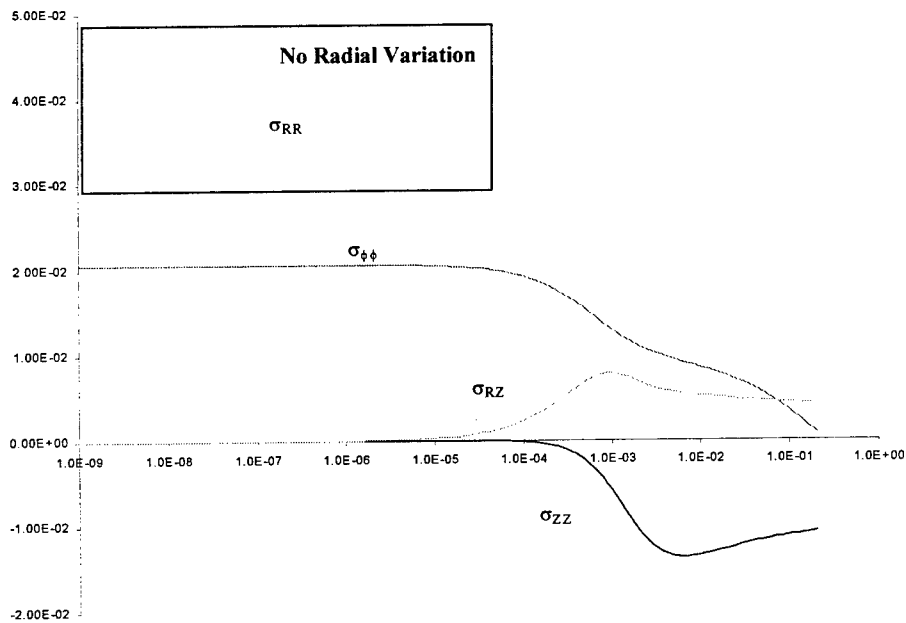**Figure 13: Variation of normal and shear stress before and after modification**

Figure 14: Variation of Axisymmetric Stress Components along the Interface (θ=90, 0 < ρ^* < Rf)

6. ACKNOWLEDGEMENTS

Thanks are due to AFOSR for providing the financial support through the summer research program and to RDL for effectively administrating the program. Author wishes to thank Dr. N. Pagano of the Materials Laboratories and Dr. G. P. Tandon of Ad Tech Systems Research for their help and collaboration during this effort. Thanks are due to Maj. Paul Hesse, Branch Chief, MLBM and all the members of the MLBM group for helpful discussions and cooperation during my stay at the Wright Laboratories. Thanks to Mr. Milton (Dan) Danishek for assistance at the base.

7. REFERENCES

1) Pagano, N.J., "On the micromechanical failure modes in a class of ideal brittle matrix composites," Wright Laboratories Report, to be published., 1997.
2) Reissner, E., "On a variational theorem in elasticity," J. Math. Phys, 29, 1950, pp. 90-95.
3) Tandon, G. P., and Pagano, N. J., "Micro-mechanical analysis of fiber push-out and re-push test," to be published, personal communication, 1997.
4) Pagano, N. J., "Axisymmetric micromechanical stress fields in composites," Proceedings 1991 IUTAM symposium on local mechanics concepts for composite material systems, Springer Verlag, 1991, pp. 1-26
5) Brown, H. W., "Analysis of axisymmetric micro-mechanical concentric cylinder model," Wright Laboratories Report.
6) Pochiraju, K. V., "Mechanics of the fiber pullout problem with several fiber-matrix interface conditions," Ph.D. dissertation, Drexel University, Philadelphia, PA, 1993
7) Pochiraju, K.V., A. C. W. Lau and A. S. D. Wang, "A local-global matching method for the single fiber pullout problem with perfectly bonded interface," Computational Mechanics, 14, 1994, pp. 84-89

APPENDIX – A: SFP REPORT - REFINED REISSNER'S VARIATIONAL SOLUTION IN THE VICINITY OF STRESS SINGULARITIES

INTRODUCTION

The ability to predict damage initiation and growth in composite structures relies on accurate stress solutions and simulation of micro-mechanical damage processes. Pagano [1] applied the approximate elasticity model based on Reissner's variational formulation [2] for the prediction of micro-mechanical failure modes in brittle matrix composites. Tandon and Pagano [3] recently investigated the failure modes and stress states of a single fiber pullout model. The solution technique, based on Reissner's variational formulation, entails dividing the structure into regions, assuming the distribution of some components of a stress field within a region, determining the spatial structure of the other components that satisfy the equilibrium and boundary conditions, and inter-region continuity conditions. Pagano [4], Kurtz and Pagano [5], and Brown [6] describe the extension of this technique to axisymmetric concentric cylinder model .

The single fiber axisymmetric model was also investigated earlier by Pochiraju [7] using a local asymptotic and a global finite element analysis. This effort employs a local-global matching method to effectively determine the detail spatial structure and magnitude of the locally singular stress field as well as the complete global stress distribution in the single fiber concentric cylinder model. The local-global matching method consists of three components: a local analysis, a global analysis, and proper matching of the local asymptotic field to the global complete stress field. (Pochiraju, Lau, Wang [8]). The local regions are identified as the areas where the geometry and the presence of interface leads to stress accentuation and possibly a singularity. In such local regions, asymptotic solutions are obtained for several fiber-matrix interface conditions. In order to determine the effects of applied thermal residual loading, mechanical pulling and boundary conditions on the local fields, a global finite element analysis is performed with mesh refinement in the local regions. No singularity information is embedded into the finite element analysis. Matching the angular variation and radial variation of stress fields obtained from the finite element analysis and with those obtained from asymptotic analysis determines the nature of the local stress fields in the local regions. The method as developed is applicable for various fiber-matrix interface conditions: namely perfectly bonded interface, partially debonded interface with interfacial crack, or debonded interface with frictional interface sliding [9].

The motivation to solve for the local stress field is that these accentuated stresses, strains and energies are likely to induce damage. The research presented in this report is motivated to include the characteristics of asymptotically singular stress fields within the Reissner's variational solution. There are several significant differences in the nature of the variational and asymptotic solutions and are listed below:

- The variational solution solves the axisymmetric equilibrium, boundary conditions and the constitutive relationships. Separable solutions with power type of radial singularities exist for plane strain problems. However, Zak [10] showed, that through the transformation of coordinates from global to local systems, that the axisymmetric solution degenerates (curvature terms become insignificant) to the plane strain solutions in small local regions around a point.
- The variational solution for axisymmetric problems is separable in the axial (Z) and radial (R) directions. The asymptotic stress fields are the solutions separable in the polar coordinate system (ρ, θ) with the origin at the point of singularity.
- The variational solution is determined by assuming the radial (R) distribution of two of the four stress components and determining the functional form of the other components by the minimization of the potential energy. This is a complete solution and satisfies the equilibrium and boundary conditions. The asymptotic solution is a partial solution and consists of undetermined coefficients. These coefficients may be obtained from loading and boundary conditions or by comparison with a global complete solution.

The variational formulation, as implemented in the Axisymmetric Damage Model, assumes a linear radial variation of the axial and hoop stresses within any solution element (Layer) and the axial variation is determined as a solution to a first order differential equation. It may be conceptualized that an assumption of stress fields of the type given in Eq. (1) for M singular fields and with N terms/field expansion of asymptotic solution for two stress components and deriving the nature of the other components using the variational technique, but such an effort would be algebraically tedious, requiring lengthy development for concept validation.

$$\sigma_{\vartheta\vartheta} = \sum_{J=1}^{M}\left(\sum_{i=1}^{N_j} K_i \rho_j^{\lambda_i}\left\langle\widetilde{\sigma}_{\vartheta\vartheta}(\theta_j)\right\rangle\right), \sigma_{zz} = \sum_{J=1}^{M}\left(\sum_{i=1}^{N_j} K_i \rho_j^{\lambda_i}\left\langle\widetilde{\sigma}_{zz}(\theta_j)\right\rangle\right) \tag{1}$$

Methods for enhancing the solutions near bi-material cracks and corners by selectively using the asymptotic solution are investigated.

COMPARISON OF ASYMPTOTIC AND VARIATIONAL SOLUTIONS

In this section, the asymptotic and axisymmetric variational solutions are compared for several crack and corner problems. Figure 1 shows schematics of the concentric cylinder model along with the R-Z meridian plane for the axisymmetric problem, and the local solution for a typical fiber pullout problem. A local region is shown in the detail-A which surrounds the point of singularity, O. First a short description of the asymptotic solutions in the local region along with a method of matching the local solution with a global numerical solution is presented. The variational solution is summarized in the second sub-section. An algorithm for determining the angular variation of stresses around the singular point from the variational solution is presented. A series of problems involving cracks and corner singularities are investigated. A comparison angular variation of the stress distributions obtained from the variational solution and asymptotic analyses is presented. This comparison is conducted at a selected radius where the leading singular term in the asymptotic analysis is believed to be dominant. A second global solution from Finite Element Method (FEM) is also presented for reference.

Local Asymptotic Solution

The point, O, is the location of material and geometric discontinuity. The local stresses at O may be singular in nature depending upon the fiber and matrix properties and the fiber-matrix interface conditions. In order to determine the spatial distributions of the stress fields in the local region, the governing equations are formulated for axisymmetric stresses in {R, Z, Φ} coordinates and then transformed to the local {ρ, θ} coordinates, retaining only terms valid as $\rho \to 0$. Stress fields separable in the ρ & θ coordinates and with radial power-type of singularities are considered. Eq. (2) shows the function form of stress components near the point of singularity as determined from the asymptotic analysis.

$$\sigma_{ij}(\rho,\theta) = k\rho^{\lambda}\widetilde{\sigma}_{ij}(\theta) \quad (\lambda < 0; \rho \to 0) \tag{2}$$

The angular distribution functions, $\widetilde{\sigma}_{ij}(\theta)$, for the matrix-domain and fiber-domain are determined from the Williams[11] type of stress functions. Eq. (3) shows the typical stress function required to derive the stress components for axisymmetric problem.

$$\varphi(\theta) = A\cos(\lambda+3)\theta + B\sin(\lambda+3)\theta + C\cos(\lambda)\theta + D\sin(\lambda)\theta$$
(3)

The traction free boundary conditions at $\theta=0$ and $\theta=3\pi/2$, and the perfectly-bonded fiber-matrix interface conditions at $\theta=\pi/2$, form a system of eight homogeneous equations ([7]):

$$[C_{ij}(\lambda)]\, A = 0 \tag{4}$$

where λ is the eigenvalue, and $\mathbf{A} = [A^m, B^m, C^m, D^m, A^f, B^f, C^f, D^f]^T$ is its associated eigenvector. For nontrivial solution, the determinant of $[C_{mn}(\lambda)]$ must vanish. The eigenpair $\{\lambda, \mathbf{A}\}$ determined from Eq. (4) provides the ρ- and θ-dependent structure, ρ^λ and $\tilde{\sigma}_{ij}(\theta)$, of the singular field. A determinant search algorithm was used to find the eigenvalue λ within the admissible range of $0 > \lambda > -1$. For some fiber-matrix combinations, there can be more than one solution for λ in the admissible range. We are concerned with the smallest admissible λ, which gives rise to the most dominant singular stresses. The θ-dependence of the singular field, $\tilde{\sigma}_{ij}(\theta)$, is determinable from solution of the eigenvector $\mathbf{A}$ to within an arbitrary constant. We normalized $\tilde{\sigma}_{ij}(\theta)$ by choosing the arbitrary constant so that $\tilde{\sigma}_{\theta\theta}(\theta^*) = 1$, where θ^* is the angular position of points on the interface.

The amplitude factor k of Eq. (2) is to be determined from the full-field solution. This is accomplished by appropriately matching the singular field with the full-field stress distribution provided by an analyses methods such as FEM or other variational methods. This matching was carried out earlier [8,9] using the FEM. A displacement-based axisymmetric finite element model, as shown in Figure 16, with quadratic Lagrange elements was used to compute the global full-field solution. Near the fiber-protrusion point O, elements were arranged in concentric rings with boundaries given by: $(\rho/R_o)_n = 10^{-8} \{1 + (10^{0.2})^{n-1}\}$ where n denotes the n-th ring. The first ring encircling the fiber-protrusion point consisted of quadratic triangular elements degenerated from Lagrange quadrilateral elements which formed the rest of the mesh. No special implementation for singularity was imposed on these elements. Within the finite element spatial domain, we searched for a sub-region where the full-field stresses, $\sigma_{ij}(r,\theta)^{FEM}$, have the identical θ–dependence as that of the singular solution, $\tilde{\sigma}_{ij}(\theta)$. If the finite element stresses match $\tilde{\sigma}_{ij}(\theta)$, then they also match the eigenvector $\mathbf{A}$ and its associated eigenvalue λ. One may choose any stress component, say $\sigma_{\theta\theta}$, at any angle, say θ^*, to determine k:

$$ k = \frac{\sigma_{\theta\theta}\left(\rho,\theta^*\right)^{FEM}}{\rho^\lambda \tilde{\sigma}_{\theta\theta}\left(\theta^*\right)} \tag{5} $$

However, starting from the second ring of elements, results show good matching characteristics typical of Figure 16. It is expected that as ρ/R_o increases, the one-term singular solution will deviate from the full-field solution, resulting in less spectacular angular matching at sampling points with large ρ/R_o. A root-mean-square error parameter, E, was developed to quantify mismatching along any arc of constant radius ρ. Finite element results were sampled at N points of varying θ-value along the arc. At each sampling point p (at θ_p), the average of the squares of the mismatch for three stress components. The expression for the error measure is shown in Eq.(6). E is zero when there is a perfect match. K is determined from the regions where E is minimum.

$$E = \sqrt{\frac{\sum_{p=1}^{N} Average\left\{\left[\frac{\sigma_{ij}(\rho,\theta_p)^{FEM}}{k\rho^\lambda} - \widetilde{\sigma}_{ij}(\theta_p)^{asymp}\right]^2\right\}_p}{N}}$$ (6)

Variational Solution

The axisymmetric damage model was developed in order to approximate the elastic stress field and energy release rates of bodies in the form of concentric cylinders and contains damaged regions in the form of annular or penny-shaped cracks in the constituents and/or debonds between them. The model is generated by subdividing the body into regions consisting of a core and a number of shells of constant length and satisfying the Reissner variational equation with an assumed stress field in each region. That is, we set

$$\delta J = 0$$ (7)

$$J = \int_V F dV - \int_{S'} \widetilde{T}_i U_i ds$$ (8)

$$F = \frac{1}{2}\tau_{ij}\left(U_{i,j} + U_{j,i}\right) - W\left(\tau_{ij}, e_{ij}\right)$$ (9)

In these equations, W is the complementary energy, τ_{ij} and U_i are the stress and displacement components, respectively, in Cartesian coordinates, e_{ij} are the mathematical free expansional or non-mechanical strains, V is the volume enclosed by S, S' is the part of S on which one or more traction components are prescribed and $\widetilde{T}_i$ are the Cartesian components of the prescribed tractions. Further, a comma followed by a subscript(s) implies differentiation with respect to the appropriate coordinate(s) and the summation convention is understood.

Reissner (1950) has shown that the governing equations of elasticity can be obtained as a consequence of the variational equation provided both stresses and displacements are subject to variation in the application of (7). The stress field is assumed such that σ_{ZZ} and $\sigma_{\theta\theta}$ are linear in R within each region, while the forms of σ_{RR} and σ_{RZ} are chosen to satisfy the axisymmetric equilibrium equations of linear elasticity. Letting,

$$\sigma_1 = \sigma_{zz}, \sigma_1 = \sigma_{\phi\phi}, \sigma_1 = \sigma_{RR}, \sigma_1 = \sigma_{RZ}$$ (10)

(and the analogous relation for the engineering strain components ε_i (i=1, 2, 3, 5) and hygrothermal free expansional strain components e_i (i = 1, 2 3)), we arrive at the relations in the region $R_1 \le R \le R_2$:

$$\sigma_i = p_{ij} f_j^{(i)} \quad (\ i = 1,2,3,5; \ J=1,2,\ldots5)$$ (11)

where $f_j^{(i)}$ are known shape functions of R defined such that $p_{i\alpha} = \sigma_i(R_\alpha, Z)$ (i = 1,2,3,5; J=1,2,...5) (12)

In order to maintain consistency with our postulated stress field, we further assume that eqs. (11) hold on the boundaries, as well as within the medium itself. The remaining dependent variables follow directly from the mathematics without further assumptions. They are the weighted displacements:

$$(\overline{u}, u^*, \hat{u}, \overline{\overline{u}}) = \int_{r_1}^{r_2} u(1, r, r^2, r^3) dr \tag{13}$$

$$(w^*, \hat{w}) = \int_{r_1}^{r_2} w(r, r^2) dr \tag{14}$$

The interfacial displacements u_α, w_α ($\alpha = 1, 2$), or displacements on the radial boundaries $R = r_1, r_2$, only enter the formulation if they are prescribed or if we have relative slip with friction at the interface. For a body composed of a core region plus N shells, the formulation leads to the solution of $18 N + 16$ algebraic and ordinary differential equations in Z. On planes, Z=constant, the weighted displacement or traction quantities are specified in certain pairs. Furthermore, continuity conditions can be written for the surfaces R = constant which are internal to the medium. Finally, the boundary conditions on $z = z_1, z_2$, or <u>end conditions</u>, can be expressed by prescribing one term from each of the following products The regions are selected such that the thermoelastic properties are constant and the boundary conditions do not change character on any of the bounding surfaces (i.e., traction and/or displacement components prescribed or continuity prescribed) within each region. The number of regions, in particular in the r direction, can be increased in order to improve the solution accuracy.

Since the field equations within each material are linear differential equations with constant coefficients, the general form of the solution for any of the dependent variables $P(z)$ is expressed by:

$$P(Z) = P_h(Z) + \sum_i A e^{\lambda_i Z} + P_p(Z) \tag{15}$$

within <u>each</u> layer where A_i are constants, λ_i are eigenvalues of a determinant, $P_h(z)$ is the polynomial type of homogenous solution (repeated roots), and $P_p(z)$ is a particular solution, which in the present case is a simple polynomial. This completes the boundary value problem formulation with the variational model.

Angular variation determination from variational solution

To tabulate the stress fields in the ADM solution in a region, following algorithm is used.

1. Given the point of singularity, R_0 and Z_0
2. Given ρ, generate a series of θ values from θ_1 start to θ_2 at $\Delta\theta$ increment
3. For every ρ and θ:
    4. Determine the global coordinate R and Z
        $$R = R_0 + \rho \cos(\theta)$$
        $$Z = Z_0 - \rho \sin(\theta)$$
    5. Determine the section number, *isec*, based on the Z value
    6. Determine the section properties for isec
    7. Determine the layer number, *ilayer*, based on the R value
    8. Do a transformation of Z to local coordinate, Z_{local} of the section and layer.
    9. Construct the solution for the unknown quantities [4] for that layer at $Z = Z_{local}$
    10. Store the solution quantities and their derivatives
11. Do radial interpolation using the appropriate equations for core and coating:

For core:

$$\sigma_{ZZ}(r) = \sigma_{ZZ}(r_1) \frac{(r_2 - r_1)}{(r_1 - r_2)} + \sigma_{ZZ}(r_2) \frac{(r - r_1)}{(r_1 - r_2)}$$

$$\sigma_{\theta\theta}(r) = \sigma_{\theta\theta}(r_1)\frac{(r_2-r_1)}{(r_1-r_2)} + \sigma_{\theta\theta}(r_2)\frac{(r-r_1)}{(r1-r_2)}$$

$$\sigma_{RZ}(r) = \sigma_{RZ}(r_2)\frac{(r-r_1)}{(r_1-r_2)} + \sigma_{RZ}^{I}(r - r_2)r$$

$$\sigma_{RR}(r) = \sigma_{RR}(r_1)\frac{(r_2-r_1)}{(r_1-r_2)} + \sigma_{RR}(r_2)\frac{(r-r_1)}{(r_1-r_2)} + \sigma_{RR}^{I}(r^2-r_2^2)r + \sigma_{RR}^{II}(r^2-r_2)r$$

For Coating:

$$\sigma_{ZZ}(r) = \sigma_{ZZ}(r_1)\frac{(r_2-r_1)}{(r_1-r_2)} + \sigma_{ZZ}(r_2)\frac{(r-r_1)}{(r_1-r_2)}$$

$$\sigma_{\theta\theta}(r) = \sigma_{\theta\theta}(r_1)\frac{(r_2-r_1)}{(r_1-r_2)} + \sigma_{\theta\theta}(r_2)\frac{(r-r_1)}{(r1-r_2)}$$

$$\sigma_{RR}(r) = \sigma_{RR}(r_1)\frac{(r_2-r_1)}{(r_1-r_2)} + \sigma_{RR}(r_2)\frac{(r-r_1)}{(r_1-r_2)} + \sigma_{RR}^{I}(r^2-r_2^2)(r^3 - r_1^3 r_2^3 r + r_1 r_2(r_1+r_2))$$

$$+ \sigma_{RR}^{II}(r^2 - (r_1+r_2)r + r_1 r_2) + \sigma_{RR}^{III}\frac{(r^2-(r_1+r_2))}{r_1 r_2 r}$$

$$\sigma_{RZ}(r) = \sigma_{RZ}(r_1)\frac{r}{r_1} + \sigma_{RZ}(\frac{r}{r_2})\frac{(r-r_1)}{(r_1-r_2)} + \sigma_{RZ}^{I}\left\{(r_1+r_2)r - \frac{(r_1^2 + r_1 r_2 + r_2)r}{r_1^2 + r_2^2} + \frac{1}{r}\right\}$$

where σ^{I}, σ^{II}, and σ^{III} correspond to the solution quantities as given in Eq. (11) [4].

12. Increment θ by $\Delta\theta$
13. Check if $\theta = \theta_2$, YES: Stop, NO: goto 3

Table 2: Material properties and power of singularity for configuration in Problem-A

Material	E_f	v_f	E_m	v_m	λ
Polyester/Epoxy	3.25 GPa	0.35	4.0 GPa	0.35	−0.00463
Silicon Carbide/ Borosilicate	410 GPa	0.2	63 GPa	0.2	−0.11206

Comparison of ADM and Asymptotic Solutions

The algorithm presented in the previous section was implemented into the current axisymetric damage model. The results presented in this section are for unmodified variational solution. Several problem involving crack and corner geometry were then investigated to determine the radial and angular distribution of the stresses predicted by the variational model. In this report, the following four case studies are presented.

Problem A: Corner singularity in the single fiber push-out model: Figure 17 shows the geometry of the axisymmetric concentric cylinder model typically employed to simulate fiber push-in experiments. The singularity at the ends of the model is material dependent is given for two material combinations in Table 1.

Problem B: The problem of an interface crack with the debonded faces sliding with friction is considered. The local coordinate system at the crack tip and the geometry of the model is shown in Figure 18. The fiber and matrix materials are taken to be Polyester and Epoxy, respectively. The coefficient of friction is

assumed to be 0.52. These values correspond to recent push-in tests conducted by Bechel [12]. The power of the stress singularity for this case is −0.49605.

Problem C: The penny shaped crack in a fiber is considered as the third example. For this case the stress singularity is −0.5 and is material independent. The mode-1 loading case is considered and for this case the exact stress intensity factor for the infinite body is also known. Figure 19 shows the local coordinate system.

Problem D: The problem of a bi-material crack with open, traction-free, crack faces is considered. This is the case where the stress singularity is complex-valued. (λ=−0.5 ± 0.0076 i) Again polyester-epoxy system is considered. The local coordinate system is shown in Figure 20.

The comparison of angular variation for problem A reveals the nature of the variational solution around the corner singularities. The asympotic behavior of the opening stress ($\sigma_{\theta\theta}$) and the shear stress ($\sigma_{\rho\theta}$) are captured by the variational solution. Figure 21 and Figure 22 show the comparison of the variational and asympotic solution for problem A. The comparison is conducted at a distance of 1 μm from the singular point and the variational solution shows a good correlation with the field obtained from asymptotic analysis. Figure 23 shows the angular distribution of the radial stress (corresponds to the axial stress σ_{ZZ} in the axisymmetric coordinate system) for asymptotic and variational solutions. The variation of the radial stress (σ_{rr}) significantly differs from the asymptotic behavior. The variational solution tends to a zero value at the interface consistent with the boundary condition imposed at that point in the solution. Figure 24 shows the radial variation of the axial stress component (σ_{ZZ}) in the fiber and matrix. This plot compares the stress distributions obtained from the FEM and variational solutions. There is a good correlation of the stress distributions between the two solutions in regions away from the singular point (at z = 0). This behavior is due to the boundary conditions imposed in the variational solution which require that the axial stress component (σ_{ZZ}) should vanish in both fiber and matrix at z = 0.

Problem-B is the frictional crack problem. The angular variation matching is conducted comparing asymptotic (Comninou-type of crack [13]), FEM and variational solutions. Again the opening stress ($\sigma_{\theta\theta}$) and the shear stress ($\sigma_{\rho\theta}$) are compared. The comparison shows additional characteristics of the variational solution. The opening stress, Figure 25, shows a good match between all the three cases though the FEM analysis is an incremental analysis and the variational and asymptotic analyses are independent of the load history. The shear stress, Figure 26 , shows that the overall distribution of the stress component in the vicinity of the stress singularity is comparable, the discontinuities in the shear stress component in the ADM solutions correspond to the jump between the sections where shear stress continuity is not imposed.

The penny shaped crack in a polyester fiber is considered. The applied loading was pure mode-I and the exact stress intensity factor is known. Figure 27 shows the comparison of the opening stresses between asymptotic and ADM solutions. There is a good correlation between the ADM and asymptotic solutions. The shear stress angular distribution is shown in Figure 28. Once again the jump across the sections in the variational model causes the discontinuity of the shear stress. The results of the angular variation comparisons are presented for the bi-material interface crack problem in

Figures Figure 29 and Figure 30. The variational solution captures the opening and shear stress distributions as predicted by the asymptotic solution for this case also.

Stress intensity (Scale) factors can be obtained by single point matching method as described in the previous sub-section. However, since an error basis was not developed for matching using the variational solution, the radius and the component of matching is arbitrary and may be subjected to a large error. Therefore, the work on stress intensity factor determination is not presented here.

REFINING VARIATIONAL SOLUTION WITH ASYMPTOTICS

In this section a technique for embedding the asymptotic solutions within selected regions of the model is described. This technique modifies the variational solution while ensuring that the global equilibrium, boundary conditions and the constitutive relationships are satisfied and maintaining stress continuity between the asymptotic and variational solution at a selected point inside the domain.

Modified Variational Solution

Consider the axisymmetric model, corresponding to Problem-A, as shown in Figure 31. In obtaining the refined solution, a region, of radius ρ^*, is identified around the singular point and in this region the asymptotic solution determined by solving the local homogenous problem prevails and the obtained variational solution is ignored. The local asymptotic solution however is determined with unknown scale coefficient(s) which depend upon the global solution. This scale factor needs to be determined using a global solution.

Since a small region around the singular point is excluded in the variational solution, several boundary conditions are also excluded in this region. This requires additional equations so that the undetermined coefficients in the variational formulation can be computed. These additional equations are obtained by matching the asymptotic and variational solutions at selected radius

A point on the boundary between the variational and asymptotic solution is identified (ρ^*, θ^*). At this point continuity conditions are imposed such that the normal, shear and radial stresses are continuous between the variational solution and the asymptotic analysis. If there are 'M' boundary conditions associated with the singular point in the variational solution (thus determining M undetermined constants), and a 'N' term expansion of the asymptotic solution is considered, then N+M independent matching conditions are required to determine both the local and global solutions. For example in Problem-A, there are three boundary conditions associated with the singular point in the variational formulation. They are the axial stress in the fiber (σ_{zz}^f), the axial stress in the matrix (σ_{zz}^m), the shear stress (σ_{Rz}, continuous on fiber and matrix side) vanish. These conditions determine three unknown coefficients in the variational solution. If a one term asymptotic solution is considered, then one undetermined scale constant needs to be obtained by matching. Thus there are four unknown coefficients requiring four matching conditions. Following four conditions can be written at the matching radius, r*. Replace these conditions at R=Rf, Z= 0 by following conditions at R=R_F, Z= Z*

@ R=Rf & Z=Z* (Z* $\neq$ 0)

$$\sigma_{ZZ} \text{ (ADM)} = \sigma_{ZZ} \text{ (ASYMP) in fiber}$$

$$\sigma_{ZZ} \text{ (ADM)} = \sigma_{ZZ} \text{ (ASYMP) in matrix}$$
$$\sigma_{RZ} \text{ (ADM)} = \sigma_{RZ} \text{ (ASYMP)}$$
$$\sigma_{RR} \text{ (ADM)} = \sigma_{RR} \text{ (ASYMP)} \tag{16}$$

Software Implementation

The current ADM software implementation was modified to include the asymptotic behavior. Following enhancements were made to the software to include the asymptotic analysis and the refinement to the solution by embedding the asymptotic solution within the variational solution.

- A **framework** for handling several singular points within a model was implemented. Each of the singular points can have different powers of singularities and angular variations. This enables the solution modifications described in the previous section to be applied at several spatial locations simultaneously.

- A local homogenous problem can be described so that power of the stress singularity and the angular distributions can be computed within the ADM solution. The user identifies the point of singularity and describes the local homogenous boundary conditions. The **eigenvalue and eigenvectors for the local problems** are computed from within the software.

- The computation of the unknown coefficients of the solution (Eq. 15) is modified so that the boundary conditions corresponding to the singular point are removed and the equations corresponding to the matching conditions shown in Eq. (16) are formulated. The solution of the **new system of the equilibrium equations** is carried out *simultaneously* determining the unknown **coefficients in the variational solution** as well as the **scale factors** for the asymptotic fields.

- A **post-processing** enhancements to determine the modified solution by considering the presence of the asymptotic solution within a small region is also developed.

Table 3 : Asymptotic Solution

POWER OF SINGULARITY: $\lambda = -.1121$	
COEFFICIENTS OF ANGULAR VARIATION	
1	1.0000 (Arbitrarily Set)
2	1.8428
3	-1.3832
4	-1.3019
5	4.9462
6	1.0667
7	-6.2097
8	0.2440
Mixity = .1567	

Table 4: Original and modified stresses at the matching point

Stress	Original (MPa)	Modified (MPa)
σ_{ZZ}	-0.00002	-0.00835
$\sigma_{\phi\phi}$	0.01834	0.01345

σ_{RR}	0.03112	0.02339
σ_{RZ}	0.00085	0.00315

Characteristics of the refined solution

Using example problem-A, the original and modified solutions are compared. An arbitrary matching distance of $\rho^* = 1E-04$ mm is chosen for this example. The results of the local asymptotic solution is given in Table 2. This problem has a fixed mixity of normal and shear stresses and hence has one undetermined scaling constant for the local asymptotic solution.

A eleven layer variational model has 86 coefficients determined through the global equilibrium solution. In this system, three relations corresponding to the boundary conditions at the point of singularity are identified in the software as equations 18, 20 and 21. Following the method presented earlier, three equations are replaced with the matching equations given in Eq. (16). The fourth relation in Eq. (16) is written as an additional equation (87th) into the equilibrium solution. The variational equilibrium solution is solved along with the additional equation that determines the scale factor. Table 3 presents the original and modified state of stress at the matching point ($\rho^*, \theta=90$).

Figure 32 shows the axial variation of the original and modified solution plotted along the fiber-matrix interface. It can be seen that the solution away from the local singular point is unaffected and the solution in the singular region corrects towards correct asymptotic behavior. This plot also shows that the matching distance from the variational solution perspective should be larger as the influence of the boundary condition (as seen as a local maximum near the singular point in the axial and shear stresses) seem to be of the order of $\rho = 0.1$ mm. The appropriate distance for the matching is currently being investigated.

SUMMARY

In this summer project period, the characteristics of the variational solution in the vicinity of a stress singularity are investigated. Though the variational formulation does not contain any features of the asymptotic solution, the solution reflects the asymptotic nature in two of the three components. Effort was also focused on refining the solution near the stress singularities. This resulted in a technique that modifies the global equilibrium solution in the variational formulation such that an exact asymptotic solution is embedded near the point of singularity, continuity of stresses is maintained between the asymptotic solution and the variational solution. It is also shown that such modification effects only the local regions of the global solution and does not influence the solutions in the regions away from the singular points.

The modification to the variational solution presented here requires detailed investigation of several important issues.

- First is the determination of the radius, ρ^*, where both asymptotic solution and the ADM solutions are effective. In the proof of concept investigation presented here, this distance is specified arbitrarily based on the prior

experience with the angular variation matching. Identification and validation of the error measures which can be used to establish an effective radial matching distance is crucial.

- The second investigation required is to generalize the methodology several types of singular fields within the ADM model. The boundary conditions corresponding to the variational solution differs from crack problems to corner problems. Based on the findings and the development presented in this report, a general procedure of modifying the global equilibrium system of equations within in the variational formulation can be established.

- As mentioned in the report, for some problems there are several terms in the asymptotic expansion which could be singular. Considering the powers of the asymptotic expansion higher than the most dominant singular term may be required. Considering the additional terms in the asymptotic solution will increase the domain in which the solution prevails. This creates additional flexibility in accurately finding a matching radius.

- The variational solution as implemented in the ADM model considered orthotropic materials. The asymptotic expansion presented here is valid for isotropic methods. A preliminary investigation into the nature of stress fields near the interfaces of the orthotropic materials and its effect on the presented methodology may be addressed.

REFERENCES

1. Pagano, N.J., "On the micromechanical failure modes in a class of ideal brittle matrix composites," Wright Laboratories Report, to be published., 1997.
2. Reissner, E., "On a variational theorem in elasticity," J. Math. Phys, 29, 1950, pp. 90-95.
3. Tandon, G. P., and Pagano, N. J., "Micro-mechanical analysis of fiber push-out and re-push test," to be published, personal communication, 1997.
4. Pagano, N. J., "Axisymmetric micromechanical stress fields in composites," Proceedings 1991 IUTAM symposium on local mechanics concepts for composite material systems, Springer Verlag, 1991, pp. 1-26
5. Kurtz, R. D. and Pagano, N. J., Analysis of the deformation of a symmetrically loaded fiber embedded in a matrix material, Composites Engineering, 1 [1] (1991), pp. 13-27.
6. Brown, H. W., "Analysis of axisymmetric micro-mechanical concentric cylinder model," Wright Laboratories Report.
7. Pochiraju, K. V., "Mechanics of the fiber pullout problem with several fiber-matrix interface conditions," Ph.D. dissertation, Drexel University, Philadelphia, PA, 1993
8. Pochiraju, K.V., A. C. W. Lau and A. S. D. Wang, "A local-global matching method for the single fiber pullout problem with perfectly bonded interface," Computational Mechanics, 14, 1994, pp. 84-89
9. Pochiraju, K.V., A. C. W. Lau and A. S. D. Wang, "Analysis of fiber pullout or push-in with frictional sliding at the fiber-matrix interface," Composites Engineering, V 5, No. 6, 1995, pp 611-631
10. Zak, A. R., "Stress in the vicinity of boundary discontinuities in bodies of revolution", Journal of Applied Mechanics, 33 (1964), pp.150-152.
11. Williams, M. L., Stress singularities resulting from various boundary conditions at angular corners of plates in extension, Journal of Applied Mechanics, 74 (1952), pp. 526.
12. Bechel, V., & Sottos, N., "Application of debond length measurements to examine the mechanics of fiber pushout", to be published in J. Mech. Phys. Solids, 1997,
13. Comninou, M., Interface crack with friction in contact zone, Journal of Applied Mechanics, 44 (1977), pp. 780-781.
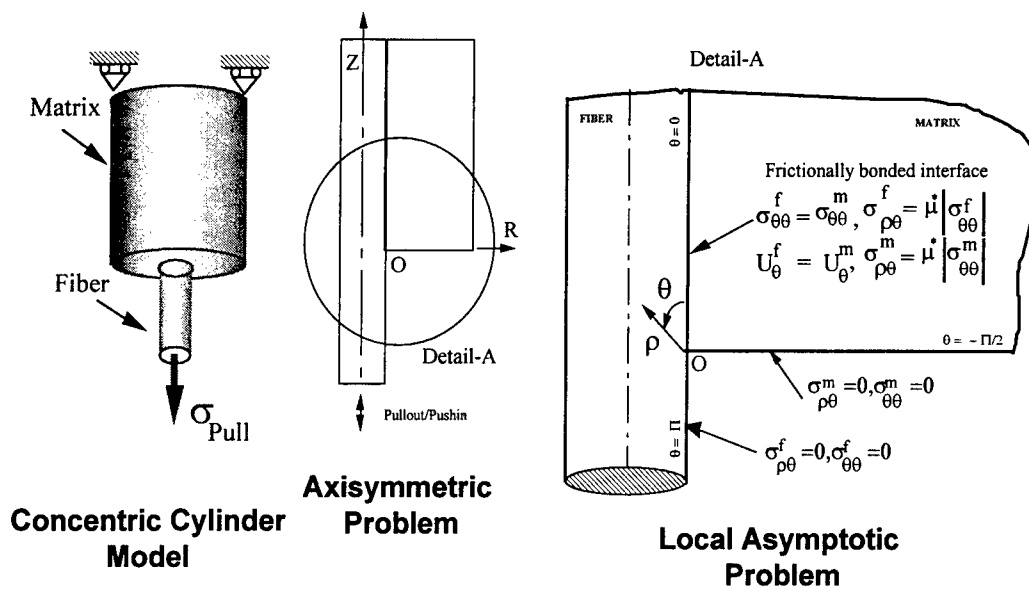
Figure 15: Concentric Cylinder Model - Global axisymmetric and local asymptotic problems
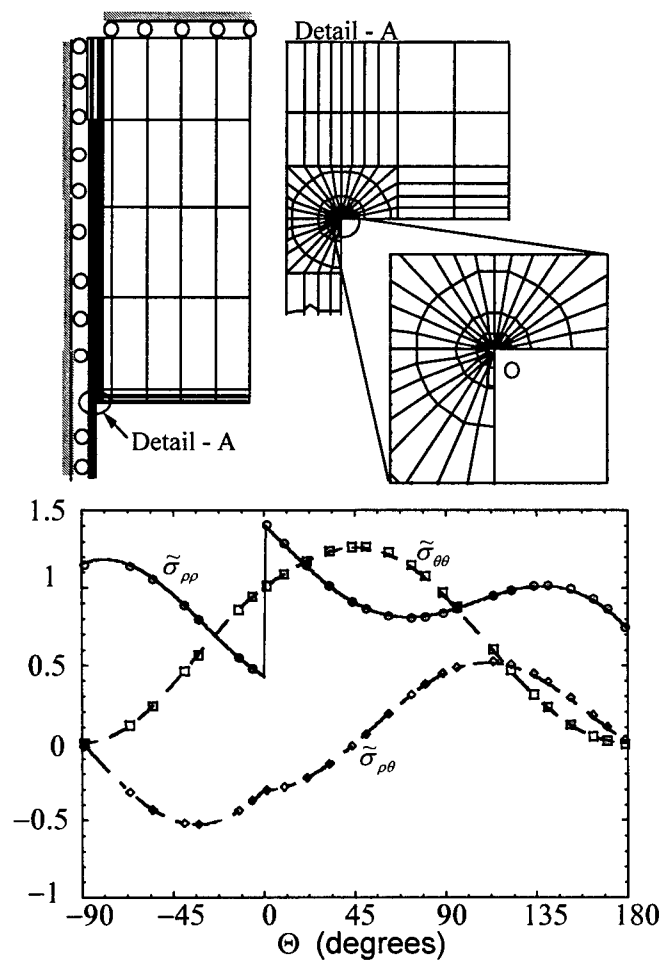

Figure 16: FEM mesh and angular variation matching.

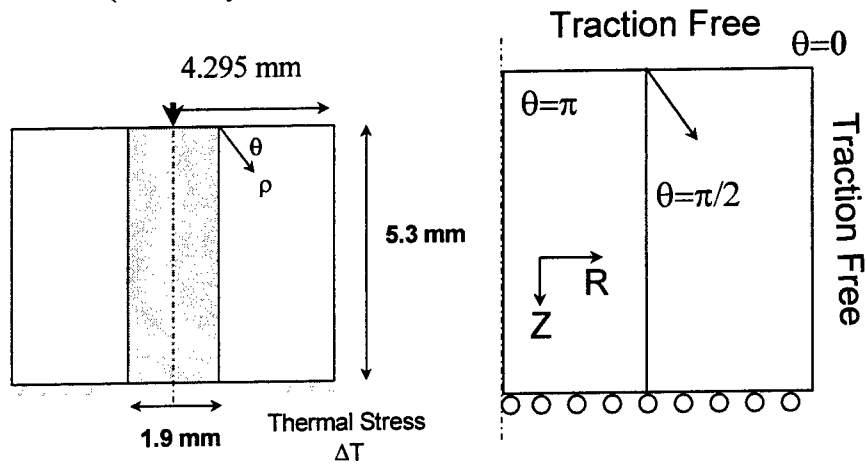**(Discrete symbols are FEM, lines are asymptotic solutions)**

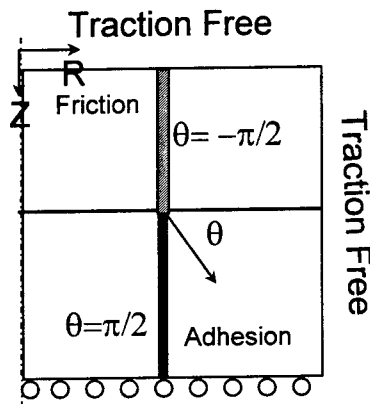**Figure 17: Problem A: Bi-material corner problem**
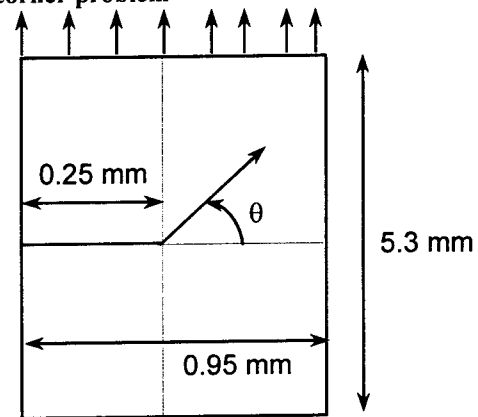

Figure 18: Problem B - Frictional Crack problem
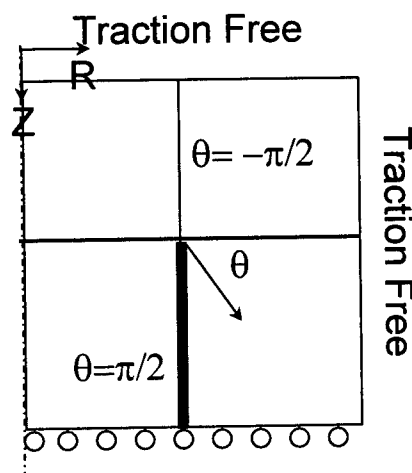

Figure 19: Problem C- Penny shaped crack

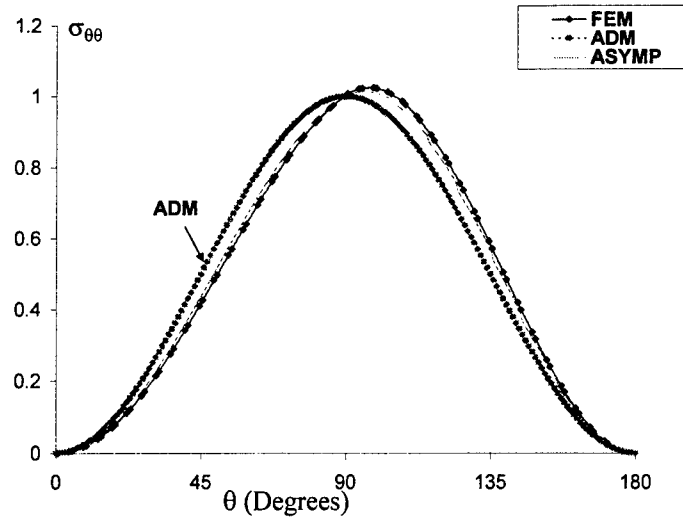**Figure 20: Problem-D: Bi-material interface crack. (Complex singularity)**

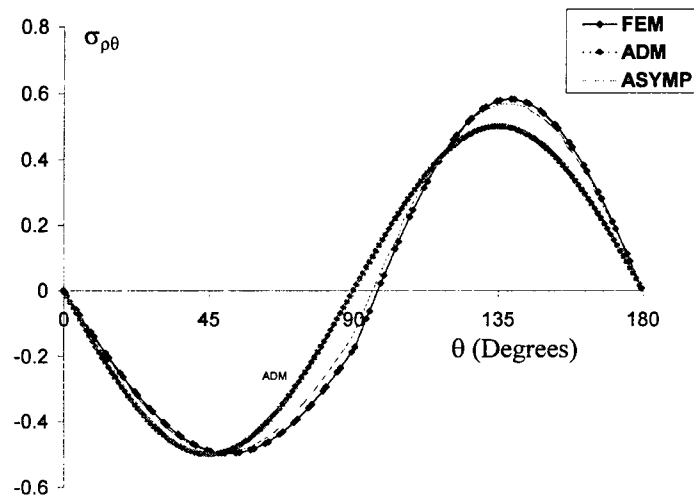**Figure 21: Opening stress angular variation comparison for problem - A**

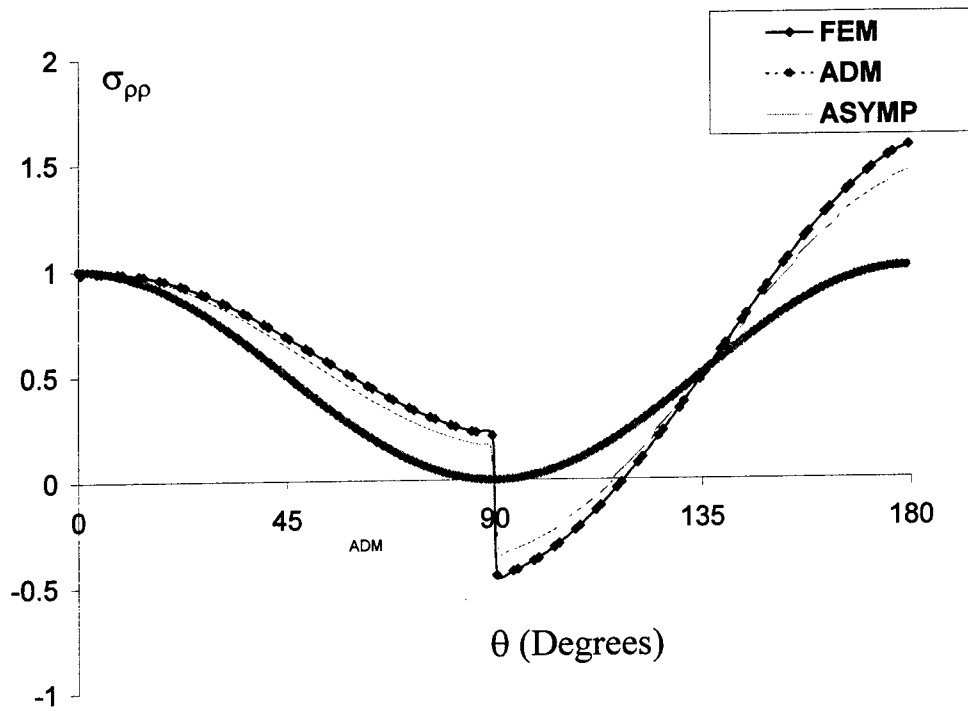**Figure 22: Shear stress angular variation comparison for problem - A**

Figure 23: Radial stress angular variation comparison for problem - A

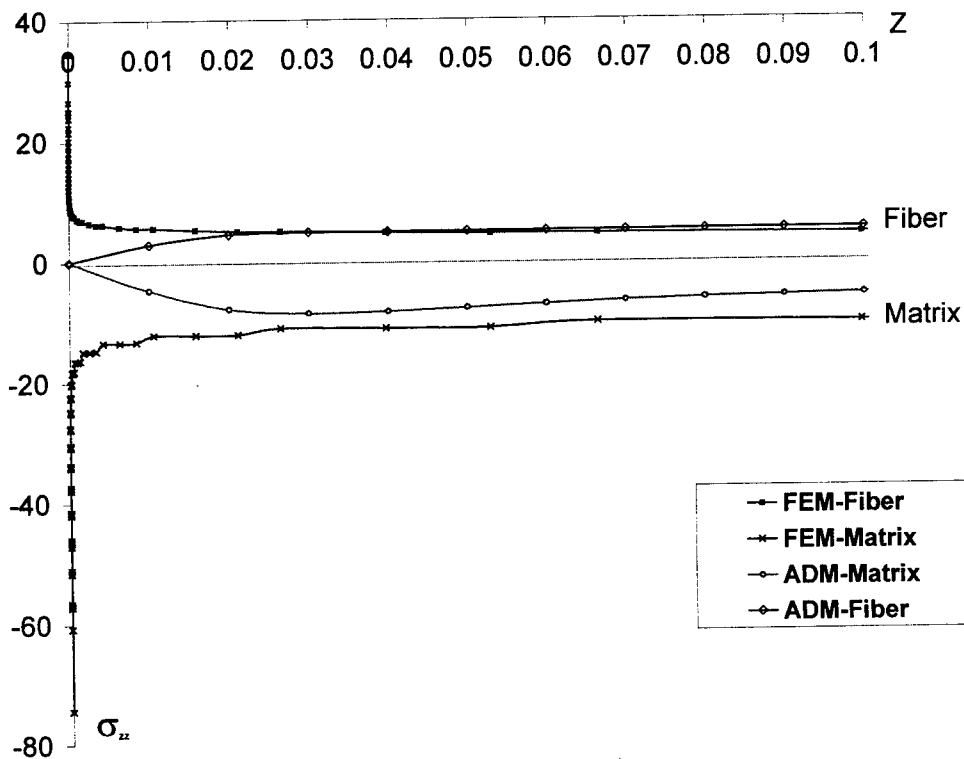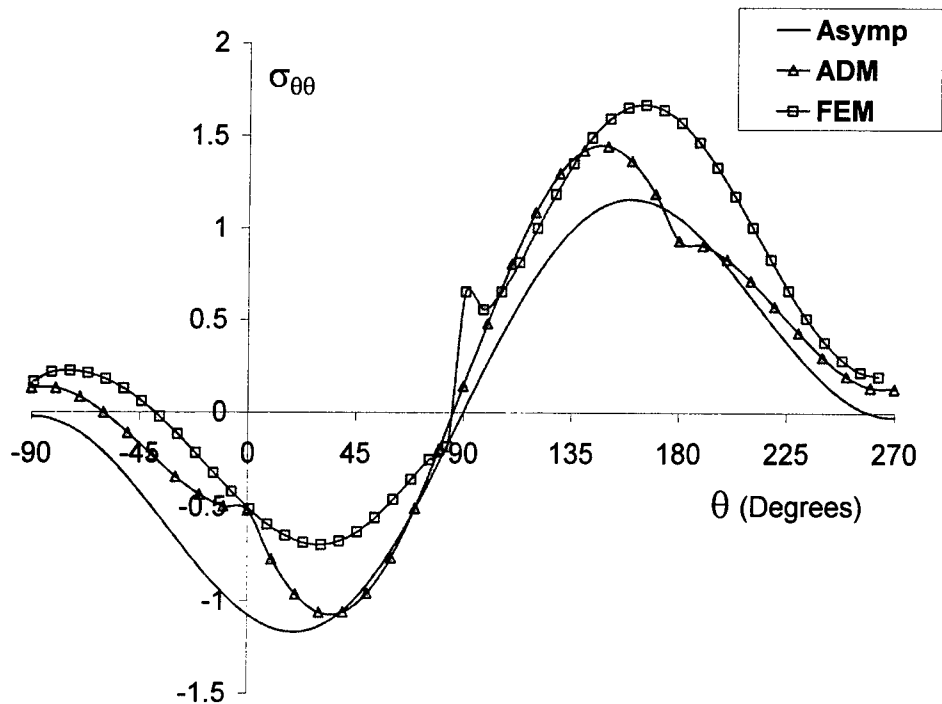**Figure 24: Radial variation of the axial stress**

Figure 25: Angular variation matching for normal stress, ADM, FEM and Asymptotic Solutions
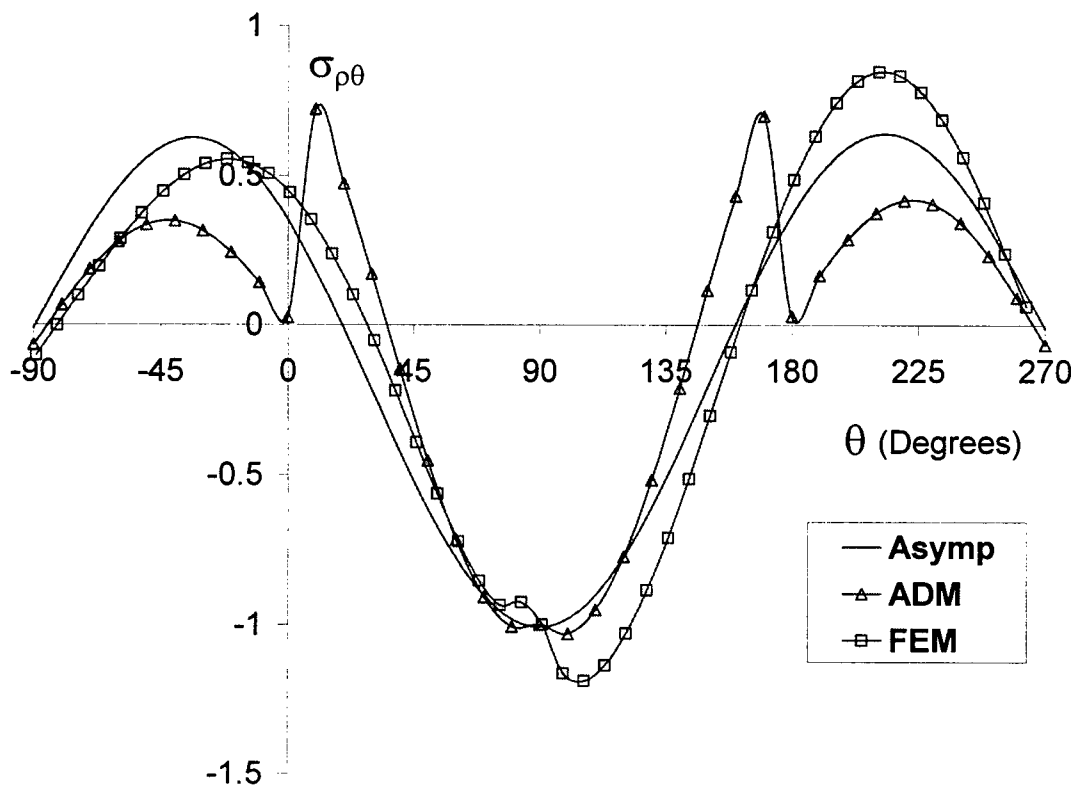

Figure 26: Angular variation matching for shear stresses ADM, FEM and Asymptotic solutions
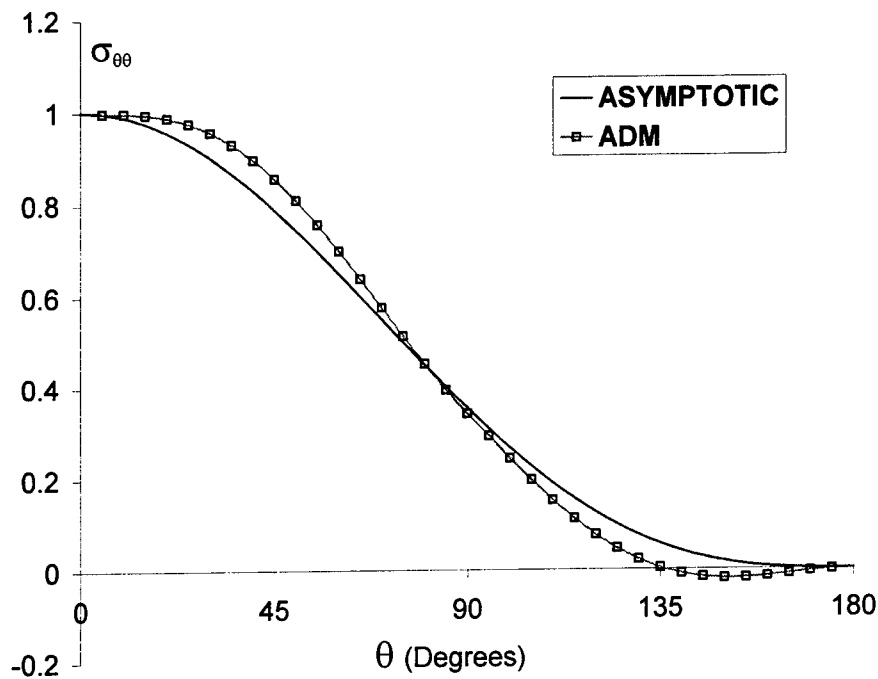
Figure 27: Angular variation matching for penny shaped crack problem (Problem-C)

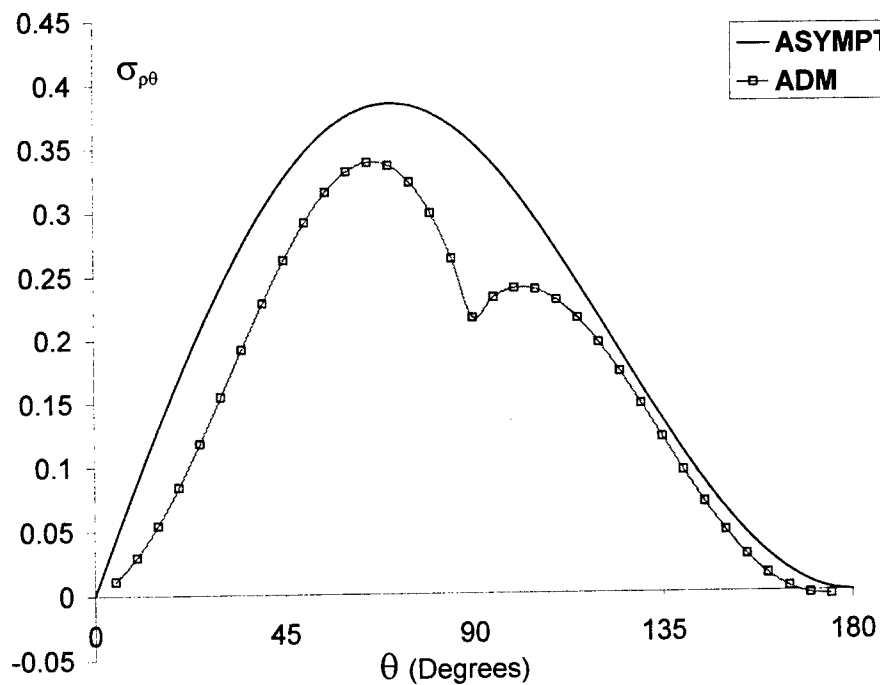

Figure 28: Angular variations of the shear stress for the penny shaped crack problem
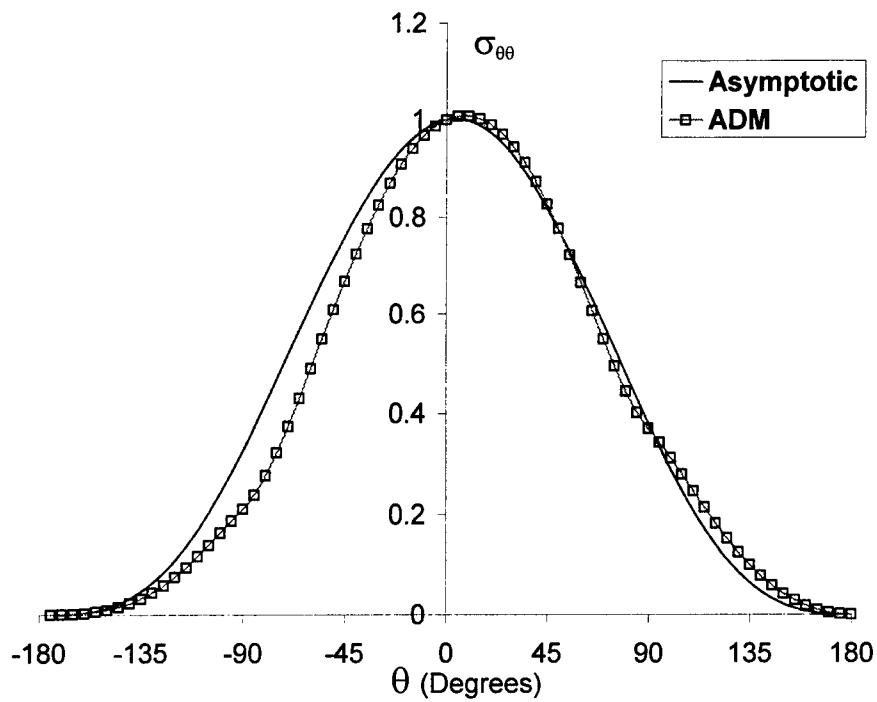
Figure 29: Normal stress distribution for the interfacial crack problem. (Problem-D)
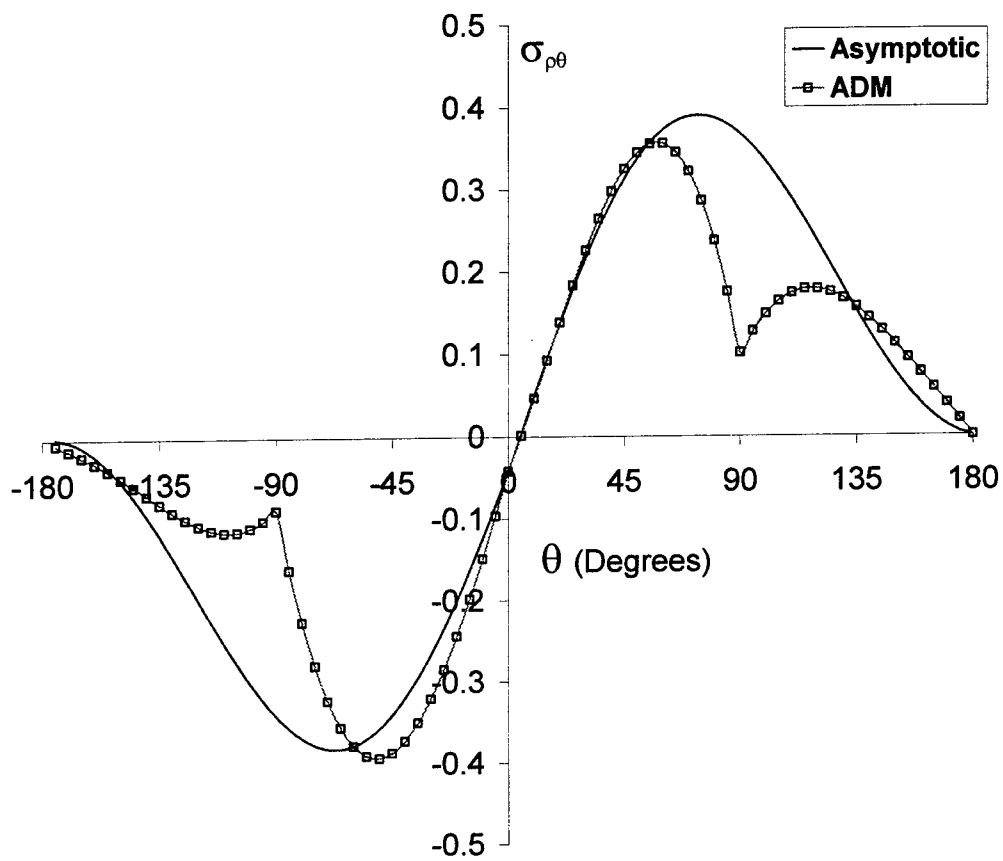
Figure 30: Shear Stress distribution comparison for problem-D.
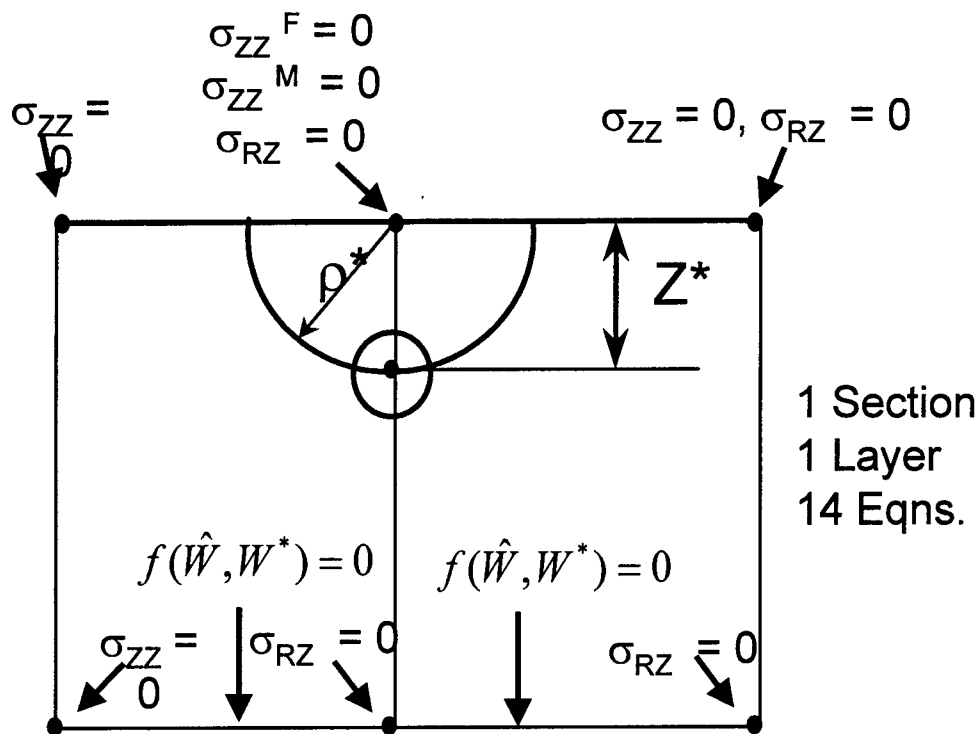
Figure 31: Refining the variational solution by embedding asymptotic solution
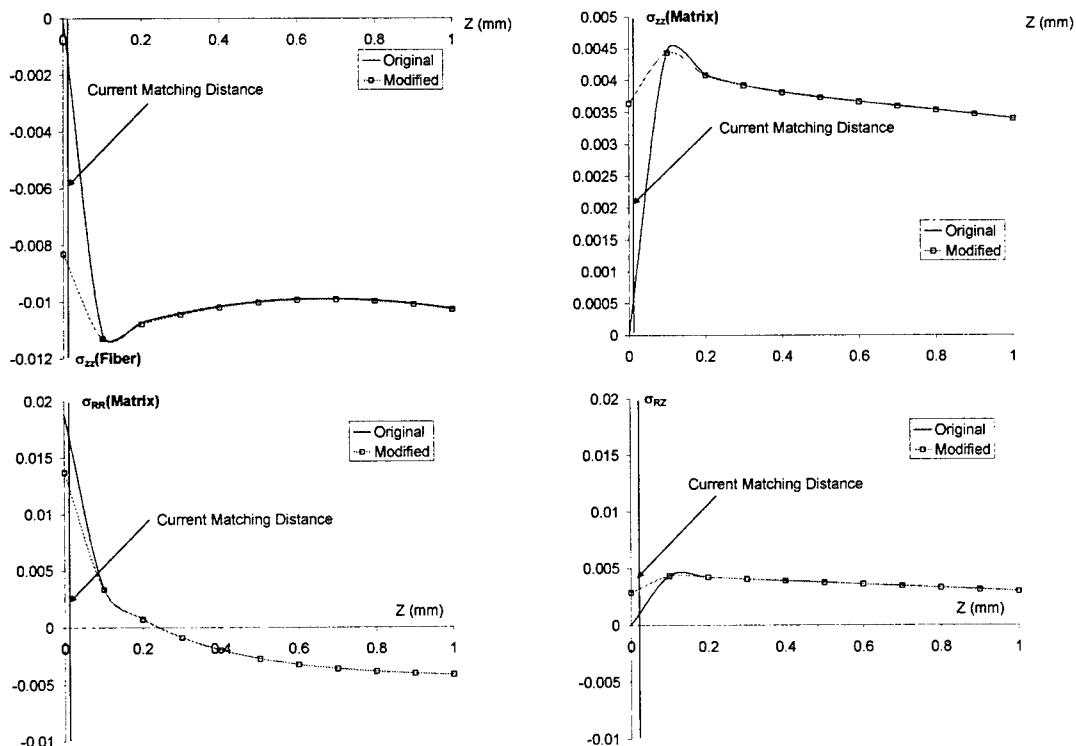
Figure 32: Original and modified solutions for problem - A

CONTINUOUS SLIDING MODE CONTROL APPROACH FOR ADDRESSING ACTUATOR DEFLECTION AND DEFLECTION RATE SARURATION IN TAILESS AIRCRAFT CONTROL AND RE-CONFIGURABLE FLIGHT CONTROL

Principal Investigator:

Dr. Yuri B. Shtessel
Associate Professor
Department of Electrical and Computer Engineering

Institution:

University of Alabama in Huntsville
301 Sparkman Drive
Huntsville, Alabama 35899

CONTINUOUS SLIDING MODE CONTROL APPROACH FOR ADDRESSING ACTUATOR DEFLECTION AND DEFLECTION RATE SARURATION IN TAILESS AIRCRAFT CONTROL AND RE-CONFIGURABLE FLIGHT CONTROL

Dr. Yuri B. Shtessel
Associate Professor
Department of Electrical and Computer Engineering
University of Alabama in Huntsville
301 Sparkman Drive
Huntsville, Alabama 35899

Abstract

A dual time scale aircraft flight control problem is addressed via continuous sliding mode control. Sliding surface boundary layer reconfiguration is used to account for actuator dynamics, deflection limits and rate limits. A re-configurable sliding mode flight controller is designed that achieves robust, high accuracy tracking of outer loop command angles before and after damage to an aircraft. Angular rate commands are robustly tracked in an inner loop. The re-configurable flight control strategy is based on a continuous sliding mode controller with direct boundary layer adaptation for re-configuration. On-line explicit system or damage identification is not required.

The continuous sliding mode controller with a finite reaching time is designed for multi-input-multi-output nonlinear control systems. Enhanced robustness issues are addressed. Computer simulations confirmed high robustness and accuracy of the designed controller. The developed controller design technique is incorporated into continuous SMC design in outer loops of traditional and tailess aircraft.

The re-configurable sliding mode flight control technique is applied to a nonlinear flight dynamics model of an F-16 aircraft and to a linearized flight dynamics model of a tailess aircraft, developed under the Innovative Control Effectors (ICE) program. Computer simulations demonstrate stability and high accuracy tracking performance without violation of actuator limits.

CONTINUOUS SLIDING MODE CONTROL APPROACH FOR ADDRESSING ACTUATOR DEFLECTION AND DEFLECTION RATE SARURATION IN TAILESS AIRCRAFT CONTROL AND RE-CONFIGURABLE FLIGHT CONTROL

Dr. Yuri B. Shtessel
Associate Professor
Department of Electrical and Computer Engineering
University of Alabama in Huntsville
301 Sparkman Drive
Huntsville, Alabama 35899

Introduction

Re-configurable flight control has been extensively investigated by the United States Air Force[1-4] due to safety and performance benefits of the technology. Recently, indirect adaptive control approaches to flight control re-configuration have been employed which supply information of system failures or damage to control law gain computation algorithms[2-9]. On-line parameter identification has been done analytically[3,4] or on the basis of a rule-based expert system[8,9]. Identified aircraft parameters are supplied to control algorithms which re-configure control laws to minimize loss of aircraft performance in the presence of control effector deflection and deflection rate limitations[2-5]. The most severe challenges of indirect adaptive flight control approaches are development of timely on-line aircraft parameter identification algorithms and directly accounting for actuator saturation. Failure to overcome these challenges may lead to significant tracking performance degradation, pilot induced oscillations and possibly loss of stability. Insensitivity and robustness of Sliding Mode Controllers (SMC) to plant disturbances and uncertainties[10-12] makes SMC attractive for flight control reconfiguration. A SMC design consists of two steps. In the first step, a sliding surface is chosen such that the closed loop system motion on this surface exhibits desired behavior regardless of plant uncertainties, and disturbances. In the second step, a control function is chosen to provide reaching of the sliding surface by the system state in finite time and guarantee system motion in this surface thereafter. System motion on the sliding surface is called a sliding mode, and strict enforcement of the sliding mode typically leads to discontinuous control functions, which may lead to control chattering[13]. However, the chattering issue is easily addressed by a continuous approximation of discontinuous control functions[13,14], or by a specially designed robust, continuous SMC[15]. Flight control re-configuration without on-line parameter identification was previously demonstrated[16,17] through angular rate tracking in the presence of damage. In this paper, a two-time scale aircraft re-configurable flight control problem is addressed via continuous SMC. Robust tracking of angle of roll, attack and sideslip angles is achieved through an outer loop SMC by providing angular rate commands to an inner loop SMC which provides robust tracking of aircraft angular rates. A sliding surface boundary layer exists for continuous SMC, and adaptation of this boundary layer is used to prevent violation of actuator bandwidth, deflection, and rate limits. The continuous sliding mode controller with a finite reaching time is designed for multi-input-multi-output nonlinear control systems. Enhanced robustness issues are addressed. Computer simulations confirmed high robustness and accuracy of the designed controller. The developed controller design technique is

incorporated into continuous SMC design in outer loops of traditional and tailess aircraft. The proposed SMC re-configurable control technique is applied to a nonlinear flight dynamics model of an F-16 aircraft and to a linearized flight dynamics model of a tailess aircraft, developed under the Innovative Control Effectors (ICE) program. Computer simulations demonstrate robust, highly accurate tracking of aggressive commands in the presence of damage without violating actuator limits.

I. Multiple Time Scale Flight Control Using Re-configurable Sliding Modes for Aircraft with Square Configuration

Problem Formulation

Consider a mathematical flight dynamics model of a rigid aircraft with the square, cascade structure

$$
\begin{cases}
\dot{y} = f_1(y,\xi) + \Delta f_1(y,\delta,\xi) + \left(B_1(y,\xi) + \Delta B_1(y,\xi)\right)z \\
\dot{z} = f_2(y,z,\xi) + \Delta f_2(y,z,\xi) + \left(B_2(y,z,\xi) + \Delta B_2(y,z,\xi)\right)\delta \\
\dot{\xi} = f_3(y,z,\xi,\delta) \\
\dot{\delta} = -A_\delta(\delta - u),
\end{cases}
\tag{1}
$$

where $y \in \Re^3$ is a vector of command angles ($y = \{\varphi,\alpha,\beta\}^T$, and φ,α,β, are roll angle, angle of attack, angle of sideslip respectively); $z \in \Re^3$ is a vector of angular rates ($z = \{p,q,r\}^T$, and p,q,r are roll, pitch and yaw rates respectively); $\delta \in \Re^3$ is a vector of control deflections ($\delta = \{\delta_a,\delta_e,\delta_r\}^T$ where $\delta_a,\delta_a,\delta_r$ are aileron, elevator, and rudder deflections respectively); $u \in \Re^3$ is a control vector containing actuator inputs, $\xi \in \Re^k$ is a vector containing all other states not contained in y, z, or δ; $f_1(y,\xi) \in \Re^3$, $f_2(y,z,\xi) \in \Re^3$, $f_3(y,z,\xi,\delta) \in \Re^k$ are differentiable vector-functions; $B_1(y,\xi) \in \Re^{3\times3}$, $B_2(y,z,\xi) \in \Re^{3\times3}$ are known matrices; $\Delta f_1 \in \Re^3$, $\Delta f_2 \in \Re^3$, $\Delta B_1 \in \Re^{3\times3}$, $\Delta B_1 \in \Re^{3\times3}$, $\Delta B_2 \in \Re^{3\times3}$ are unknown bounded smooth perturbations, caused by damage, failures and modeling uncertainties; $\det\left(B_1(y,\xi) + \Delta B_1(y,\xi)\right) \neq 0$ and $\det\left(B_2(y,z,\xi) + \Delta B_2(y,z,\xi)\right) \neq 0$ $\forall y,\xi \in \Omega$.

The deflections and deflection rates of the actuators are assumed to be bounded:

$$
|\delta_i| \leq \delta_m, |\dot{\delta}_i| \leq \overline{\delta}_m \ \forall i = \overline{1,3}.
\tag{2}
$$

Problem: Given a real-time command reference profile $y_c(t)$, design a continuous SMC to obtain the robust asymptotic de-coupled motion of the output tracking errors

$$
\lim_{t\to\infty} |y_{ci} - y_i| = 0 \ \ \forall i = \overline{1,3}
\tag{3}
$$

in sliding modes described by homogeneous time-invariant equations with given eigenvalues placement for the system in eqs. (1) and (2).

Assumption 1. It is assumed that the equations

$$
\dot{\xi} = f_3(y_c,z_c,\xi,\delta_c)
\tag{4}
$$

are stable, where δ_c is the commanded deflection that satisfies equation $\dot{\delta}_c = -A_\delta (\delta_c - \tilde{u})$, and $\tilde{u}$ is the control that maintains commanded states y_c, z_c.

This is a reasonable assumption since if any elements of ξ were not stable, it is argued that those elements should be moved to either y or z depending on the dynamical structure. This assumption is analogous to the minimum phase assumption used in feedback linearization.

Flight Control System

The cascade structure of the system in eq. (1) and the inherent two time scale nature of the proposed flight control problem are exploited for design of a two loop flight control system using continuous SMCs in the inner and outer loops. The inner loop SMC provides robust de-coupled tracking of angular rates without violating actuator limits. The outer loop SMC provides angular rate commands to the inner loop. Together, the inner and outer loop SMCs form a two loop flight control system that achieves de-coupled asymptotic tracking of the command angle reference profile. Similar structure was developed[18] for an aircraft control system using a dynamic inversion algorithm.

Outer Loop Continuous Sliding Mode Controller Design

Considering the angular rate vector z as a virtual control command z_c, only the command angle dynamics are considered for the outer loop SMC design

$$\dot{y} = f_1(y,\xi) + \Delta f_1(y,\delta,\xi) + (B_1(y,\xi) + \Delta B_1(y,\xi))z .$$ (5)

The vector relative degree of eq. (5) is $\bar{r} = [1,1,1]^T$, and the sliding surface can be designed as follows[10-13]: $\sigma = e = 0$. However, in order to have desired command angle error dynamics and to nullify a steady state tracking error due to continuous implementation of SMC the sliding surface is designed with an integral term in the PI format in the first step of the outer loop SMC design

$$\sigma = e + C^1 \int_0^t e\, d\tau = 0, \qquad \sigma, e \in R^3 ,$$ (6)

where $e = \{e_1, e_2, e_3\}^T$, $e_i = y_{ci} - y_i$ is the command angle error and $C^1 = diag\{c_i^1\}$ $\forall i = \overline{1,3}$ is chosen to achieve desired command angle error dynamics.

The second step of the outer loop SMC design is to choose a control function that achieves asymptotic stability of the σ dynamics given as

$$\dot{\sigma} = \Psi(.) - [B_1(.) + \Delta B_1(.)]z_c ,$$ (7)

where

$$\Psi(.) = \dot{y}_c - f_1(.) - \Delta f_1(.) + C^1 e .$$ (8)

Consider the following Lyapunov function candidate for assistance in the design of a SMC

$$V = \frac{1}{2}\sigma^T \sigma > 0 .$$ (9)

The derivative of the Lyapunov function candidate is

$$\dot{V} = \sigma^T \{\Psi(.) - [B_1(.) + \Delta B_1(.)]z_c\} .$$ (10)

A SMC of the following form is chosen

$$z_c = \hat{z}_{c_{eq}} + \mu B_1^{-1}(.)\,SIGN\,\sigma,$$ (11)

where $SIGN\,\sigma = col\{sign\,\sigma_i\}\ \forall i = \overline{1,3}$, μ is a positive scalar and

$$\hat{z}_{c_{eq}} = B_1^{-1}(.)\hat{\Psi}(.), \quad \hat{\Psi}(.) = \dot{y}_c - f_1(.) + C^1 e.$$ (12)

The equivalent control [10,11] $z_{c_{eq}}$ is the control that provides a zero Lyapunov function derivative similar to a feedback linearization term[20]. With the SMC in eq. (11), the derivative of the Lyapunov function is given by the following

$$\dot{V} = \sigma^T \{\Phi(.) - \mu(I + E(.))\,SIGN\,\sigma\},$$ (13)

where

$$E(.) = \Delta B_1(.) \cdot B_1^{-1}(.), \quad \Phi(.) = \Psi(.) - (I + E(.))\hat{\Psi}(.).$$ (14)

Assuming the following bounds

$$|\Phi_i(.)| \le \gamma_i, \quad |E_{ij}(.)| \le \omega_{ij}, \quad \sum_{j=1}^{3} \omega_{ij} < 1 \ \forall i = \overline{1,3}\ \forall j = \overline{1,3}$$ (15)

are satisfied, it is easy to show that

$$\dot{V} \le -\sum_{i=1}^{3} |\sigma_i| \cdot \left[\mu\left(1 - \sum_{j=1}^{3} \omega_{ij}\right) - \gamma_i\right].$$ (16)

Requiring $\dot{V} \le -\sum_{i=1}^{3} \tilde{\mu}_i |\sigma_i|$ will guarantee asymptotic origin convergence, for the system in eq. (7) via outer loop SMC (eq. (11)), in finite time. Any desired reaching time may be achieved by proper choice of $\tilde{\mu}_i > 0$ [11,13]. The following choice of μ will guarantee a negative definite Lyapunov derivative and finite reaching specified by $\tilde{\mu}_i$

$$\mu > \max_{i=1,2,3} \frac{\gamma_i + \tilde{\mu}_i}{1 - \sum_{j=1}^{3} \omega_{ij}}.$$ (17)

Remark 1. The physical meaning of the inequalities in (15) is uncertainty bounds y-dynamical system. Coordinates of the vector-function $\Phi(.) = \Psi(.) - (I + E(.))\hat{\Psi}(.)$ must be bounded-($|\Phi_i(.)| \le \gamma_i$) in the perturbed case and zero nominally ($|\Phi_i(.)| = 0$). Elements of the relative uncertainty matrix $E(.) = \Delta B_1(.) \cdot B_1^{-1}(.)$ must be bounded in the perturbed case and zero nominally ($E(.) = 0$). Further, it is obvious that $B_1(.) + \Delta B_1(.) = (I + E(.))B_1(.)$, so the conditions $\sum_{j=1}^{3} \omega_{ij} < 1$ enforce a diagonally dominant structure to the matrix $(I + E(.))$.

Therefore, the outer-loop SMC provides asymptotic stability of the σ-dynamics in the presence of bounded uncertainties provided that there is sufficient control power. However, the angular rate command provided

by the outer loop SMC (z_c) is discontinuous and thus is incapable of being tracked by the inner loop. A continuous approximation[13,14] of the outer loop SMC in eq.(11) is used to provide feasible angular rate commands to the inner loop. The discontinuous term $SIGN\sigma$ in eq.(11) is replaced by the following continuous expression

$$\frac{\sigma}{\zeta} = \left\{ \frac{\sigma_1}{\zeta_1}, \frac{\sigma_2}{\zeta_2}, \frac{\sigma_3}{\zeta_3} \right\}^T, \tag{18}$$

where ζ_i are designer-specified constants which value are to be chosen to avoid chattering[13].

Inner Loop Continuous Sliding Mode Controller Design

The angular rate and actuator dynamics are considered for the inner loop SMC design

$$\begin{cases} \dot{z} = f_2(y,z,\xi) + \Delta f_2(y,z,\xi) + \left(B_2(y,\xi) + \Delta B_2(y,\xi) \right) \delta \\ \dot{\delta} = -A_\delta (\delta - u). \end{cases} \tag{19}$$

The actuator dynamics are included so that actuator bandwidth and rate limits may be directly accounted for in the design. A continuous SMC u is designed to obtain the robust asymptotic de-coupled motion of the tracking errors

$$\lim_{t \to \infty} |z_{ci} - z_i| = 0 \quad \forall i = \overline{1,3} \tag{20}$$

in sliding modes described by homogeneous time-invariant equations with given eigenvalues placement for the system in eqs. (19) and (2). Note that the angular rate command z_c is generated by the outer loop SMC. A sliding surface that provides desired tracking dynamics is chosen in the first step of the inner loop SMC design. The vector relative degree of eq. (20) is $\bar{r} = [2, 2, 2]^T$, and the sliding surface is designed in a PID format[10-13]. This is

$$s = \dot{\eta} + C^2 \eta + C^3 \int_0^t \eta d\tau = 0, \quad s, \eta \in R^3, \tag{21}$$

where $\eta = \{\eta_1, \eta_2, \eta_3\}^T$, $\eta_i = z_{ci} - z_i$ and $C^2 = diag\{c_i^2\}$ $C^3 = diag\{c_i^3\}$ $\forall i = \overline{1,3}$ are chosen to provide the desired, de-coupled angular rate tracking dynamics in the sliding mode. An integral term is added to nullify a steady state tracking error due to continuous implementation of SMC.

Remark 2. The inner loop sliding mode dynamics in eq. (21) must be much faster than the outer-loop sliding mode dynamics to preserve sufficient time scale separation between the loops.

The second step of the inner loop SMC design is to choose a control function u that achieves asymptotic stability of the s dynamics given by the following

$$\dot{s} = F(.) - \tilde{B}_2(.)v, \qquad \tilde{B}_2(.) = B_2(.) + \Delta B_2(.) \tag{22}$$

where

$$F(.) = \ddot{z}_c + C^2 \dot{z}_c + C^3 \eta - \frac{d}{dt}(f_2 + \Delta f_2) - C^2(f_2 + \Delta f_2) - \left[\frac{d}{dt}(B_2 + \Delta B_2) + C^2(B_2 + \Delta B_2) \right] \delta, \tag{23}$$

$$v = -A_\delta (\delta - u). \tag{24}$$

Assumption 2. The control effectiveness matrix $\tilde{B}_2$ is assumed symmetric and positive definite.

Checking the positive definiteness of the control effectiveness matrix, the signs of its columns may be changed with an appropriate sign change of the corresponding control law command elements. Most conventional flight control models nominally exhibit diagonal dominance of the control effectiveness matrix. This is

$$\tilde{b}_{ii} > \sum_{\substack{j=1 \\ j \neq i}}^{3} \left| \tilde{b}_{ij} \right|, \qquad \tilde{b}_{ii} > 0 \ \forall i = \overline{1,3} . \tag{25}$$

With diagonal dominance, application of the Gershgorin theorem[19] provides sufficiency of positive definiteness. In unconventional flight control models such as aircraft with damage, the control effectiveness matrix may not be diagonal dominant. In this case, the Sylvester theorem[13,19] may be used to verify positive definiteness.

The following Lyapunov function candidate is considered for design of a stabilizing SMC

$$V = \frac{1}{2} s^T \left(\tilde{B}_2 \right)^{-1} s > 0 . \tag{26}$$

The derivative of the Lyapunov function candidate in eq. (26) is given by the following

$$\dot{V} = s^T \left(\tilde{B}_2 \right)^{-1} \dot{s} + \frac{1}{2} s^T \frac{d}{dt} \left(\tilde{B}_2 \right)^{-1} s = s^T \left[\left(\tilde{B}_2 \right)^{-1} F(.) - \frac{1}{2} \left(\tilde{B}_2 \right)^{-1} \frac{d}{dt} \left(\tilde{B}_2 \right) \left(\tilde{B}_2 \right)^{-1} s - v \right] . \tag{27}$$

The equivalent control v_{eq} is given by

$$v_{eq} = \left(\tilde{B}_2 \right)^{-1} \left[F(.) - \frac{1}{2} \frac{d}{dt} \left(\tilde{B}_2 \right) \left(\tilde{B}_2 \right)^{-1} s \right] . \tag{28}$$

The following control law is constructed to provide negative definiteness of the Lyapunov function derivative

$$v = \hat{v}_{eq} + \rho SIGN s , \tag{29}$$

where $\hat{v}_{eq} = \left(B_2 \right)^{-1} \left[\hat{F}(.) - \frac{1}{2} \frac{d}{dt} \left(B_2 \right) \left(B_2 \right)^{-1} s \right]$ is an estimate of v_{eq}, and $\hat{F}(.)$ is an estimate of $F(.)$,

$\rho = diag\{\rho_i\}, \rho_i > 0$, $SIGN s = diag\{sign s_i\}$, $\forall i = \overline{1,3}$. Note that $\hat{v}_{eq}$ is equal to v_{eq} and $\hat{F}(.)$ is equal to $F(.)$ with all the uncertain terms set identically to zero. Assuming the following estimation error bounds for the equivalent control elements are satisfied

$$\left| \Delta_{eq_i} \right| \leq L_i, \ L_i > 0 \ \forall i = \overline{1,3}, \tag{30}$$

$$\Delta_{eq} = v_{eq} - \hat{v}_{eq} = \left(\tilde{B}_2 \right)^{-1} F(.) - \left(B_2 \right)^{-1} \hat{F}(.) - \frac{1}{2} \left[\left(\tilde{B}_2 \right)^{-1} \frac{d}{dt} \left(\tilde{B}_2 \right) \left(\tilde{B}_2 \right)^{-1} - \left(B_2 \right)^{-1} \frac{d}{dt} \left(B_2 \right) \left(B_2 \right)^{-1} \right] s , \tag{31}$$

it is easy to show that

$$\dot{V} = s^T \left(\Delta_{eq} - \rho SIGN s \right) \leq - \sum_{i=1}^{3} \left| s_i \right| \left(\left| \Delta_{eq_i} \right| - \rho_i \right) . \tag{32}$$

Requiring $\dot{V} \leq - \sum_{i=1}^{3} \tilde{\rho}_i \left| s_i \right|$ will guarantee asymptotic origin convergence, for the system in eq. (22) via inner loop SMC (eq. (29)), in finite time. Any desired reaching time may be achieved by proper choice of $\tilde{\rho}_i > 0$ [11,13]. The following choice of ρ_i will guarantee a negative definite Lyapunov derivative and finite reaching specified by $\tilde{\rho}_i$

$$\rho_i > \left| \Delta_{eq_i} \right| + \tilde{\rho}_i > L_i + \tilde{\rho}_i \quad \forall i = \overline{1,3}. \tag{33}$$

Therefore, the inner-loop SMC in eqs. (29) and (33) provides asymptotic stability of the system in eq. (22) in the presence of bounded system uncertainties provided that there is sufficient control power. In other words, robust decoupled sliding modes exist characterized by a sliding surface in eq. (21) that is reached in finite time[11,13]. The elements of the control law vector are developed by solving eq.(24) for the actuator command, u, and inserting the SMC from eq.(29)

$$u_i = \delta_i + \frac{1}{a_\delta} \left(\hat{v}_{eq_i} + \rho_i \, sign \, s_i \right) \quad \forall i = \overline{1,3}. \tag{34}$$

A continuous inner loop SMC with a boundary layer is used to avoid control chattering. The discontinuous terms $sign \, s_i$ are substituted by the following continuous terms

$$sat \frac{s_i}{\varepsilon_i} = \begin{cases} 1, & if \, s_i > \varepsilon_i \\ \dfrac{s_i}{\varepsilon_i}, & if \, \left| s_i \right| \le \varepsilon_i \\ -1, & if \, s_i < -\varepsilon_i \end{cases} \tag{35}$$

which results in the following continuous inner loop SMC

$$u_i = \delta_i + \frac{1}{a_\delta} \left(\hat{v}_{eq_i} + \rho_i \, sat \frac{s_i}{\varepsilon_i} \right) \quad \forall i = \overline{1,3}. \tag{36}$$

Although the continuous SMC in eq. (36) prevents chattering, the control function does not account for actuator limits. If there is not enough control power to compensate for disturbances and uncertainties due to damage or other modeling errors, actuator deflection and deflection rate limits may be exceeded and controller integrators may windup. Exceeding actuator limits and integrator windup may lead to significant tracking degradation up to loss of stability and/or pilot induced oscillations. The actuator limitations in eq. (2) are directly addressed by adapting the boundary layer thickness. The boundary layer ε_i is selected to satisfy the following inequalities that guarantee compliance with actuator displacement and rate limits and avoidance of integrator windup:

a) $\quad \left| s_i \right| \le \varepsilon_i \quad \forall i = \overline{1,3}$ - integrator anti-windup, $\tag{37}$

b) $\quad \left| \delta_i + \dfrac{1}{a_\delta} \left(\hat{v}_{eq_i} + \rho_i \dfrac{s_i}{\varepsilon_i} \right) \right| \le \delta_m \quad \forall i = \overline{1,3}$ - actuator deflection limit compliance, $\tag{38}$

c) $\quad \left| \hat{v}_{eq_i} + \rho_i \dfrac{s_i}{\varepsilon_i} \right| \le \bar{\delta}_m \quad \forall i = \overline{1,3}$ — actuator rate limit compliance. $\tag{39}$

Integrator windup is avoided and actuator rate and deflection limits are not violated for the following choice of sliding surface boundary layers

$$\varepsilon_i = \max \left\{ \varepsilon_i^1, \varepsilon_i^2, \varepsilon_i^3 \right\} \quad \forall i = \overline{1,3}. \tag{40}$$

where

$$\varepsilon_i^1 = \left| s_i \right| + r_i^1, \quad \forall i = \overline{1,3}, \tag{41}$$

$$\varepsilon_i^2 = \frac{|s_i|\rho_i}{a_\delta \delta_m - (a_\delta \delta_i + \hat{v}_{eq_i})sign\ s_i} + r_i^2, \qquad \forall i = \overline{1,3}, \qquad (42)$$

$$\varepsilon_i^3 = \frac{|s_i|\rho_i}{\overline{\delta}_m - \hat{v}_{eq_i}sign\ s_i} + r_i^3, \qquad \forall i = \overline{1,3}, \qquad (43)$$

and $r_i^1 > 0$, $r_i^2 > 0$, $r_i^3 > 0$ $\forall i = \overline{1,3}$ are selected to avoid chattering[13].

Remark 3. The proposed re-configuration (adaptation) of the boundary layers is implemented on-line, and therefore accounts for failures and damage without explicit parameter identification.

Remark 4. If at least one of the following inequalities is satisfied

$$\hat{v}_{eq_i}signs_i > a\delta_m - a_\delta \delta_i signs_i, \qquad \left|\hat{v}_{eq_i}\right| > \overline{\delta}_m \qquad (44)$$

only changing the boundary layer thickness according to eqs. (42) and (43) cannot satisfy the conditions (38) and (39). A following approach is be used to avoid this anomaly.

Set $\hat{v}_{eq_i} = 0$ in the continuous SMC eq. (36). The following revised expressions may be used in place of eqs. (42) and (43)

$$\varepsilon_i^2 = \frac{|s_i|\rho_i}{a_\delta(\delta_m - \delta_i sign\ s_i)} + r_i^2, \qquad (45)$$

$$\varepsilon_i^3 = \frac{|s_i|\rho_i}{\overline{\delta}_m} + r_i^3. \qquad (46)$$

Robustness.

It is well-known that the continuous SMC provides finite-time convergence only to the vicinity of the sliding surface, and robustness is considered in the context of the sliding surface boundary layer and referred to as ε -robustness. Analyzing robustness of the system's tracking performance in the sliding mode we will estimate domains of attraction of sliding surfaces in inner and outer control loops. Further, the system's tracking performance will be estimated while the system moves in these domains. The domain of attraction of the system in eqs. (22) and. (36) is analyzed by considering the derivative of the Lyapunov function eq. (27)

$$\dot{V} \leq \begin{cases} -\sum_{i=1}^{3}|s_i|(L_i - \rho_i), & if\ |s_i| > \varepsilon_i \\ -\sum_{i=1}^{3}|s_i|\left(L_i - \frac{\rho_i|s_i|}{\varepsilon_i}\right) & if\ |s_i| < \varepsilon_i \end{cases}$$

The boundary layer $|s_i| = \varepsilon_i$ will be reached in finite time since $\dot{V} \leq -\sum_{i=1}^{3}\tilde{\rho}_i|s_i|$ $\forall |s_i| > \varepsilon_i$, and a solution $s(t)$ of eq. (22) is uniformly ultimately bounded[14]. Tracking performance ε -robustness in the inner loop may be proved from the following expression

$$\left|\dot{\eta}_i + c_i^2\eta + c_i^3\int_0^t \eta_i dt\right| \leq \varepsilon_i, \quad \eta_i = z_{ci} - z_i \forall i = \overline{1,3} \qquad (47)$$

in the sense that there exist[19] $g_i > 0$ and $d_i > 0$ that $|\eta_i| \leq g_i |s_i| + d_i \leq g_i \varepsilon_i + d_i$ within the sliding surface

boundary layer ($|s_i| \leq \varepsilon_i$).

The system's (1) outer loop motion in the σ-subspace in eq (7) will be effected by an extra disturbance

term $[B_1(.) + \Delta B_1(.)]\eta$ because the command z_c is tracked in the inner loop with the η-error. This is

$$\dot{\sigma} = \Psi(.) + [B_1(.) + \Delta B_1(.)]\eta - [B_1(.) + \Delta B_1(.)]z_c. \tag{48}$$

The domain of attraction of the origin of the system in eq. (48) with the outer loop continuous SMC in eq.

(11) with a discontinuous term $SIGN\sigma$ replaced by the continuous term in eq. (18) is analyzed by considering the

derivative of the Lyapunov function in eq. (9)

$$\dot{V} = \sigma^T \{\Psi(.) + [B_1(.) + \Delta B_1(.)]\eta - [B_1(.) + \Delta B_1(.)]z_c\}. \tag{49}$$

The conditions in eq. (15) imply the positive definiteness of the matrix $(I + E(.))$. Then the following inequality is

valid[19]: $\sigma^T (I + E(.))\sigma \geq \lambda_{\min} \|\sigma\|^2$, where $\lambda_{\min}$ is a minimal eigenvalue of the matrix $(I + E(.))$. Taking into

account $\overline{\Phi}(.) = \Psi(.) + [B_1(.) + \Delta B_1(.)]\eta - (I + E(.))\hat{\Psi}(.)$ and assuming $|\overline{\Phi}_i(.)| \leq \overline{\gamma}_i(\varepsilon_i)$ and $\zeta_1 = \zeta_2 = \zeta_3 = \zeta_0$ the

continuous outer loop SMC law is substituted into eq. (49). This is

$$\dot{V} \leq -\sum_{i=1}^{3} |\sigma_i| \cdot \left[\lambda_{\min} \frac{\mu}{\zeta_0} |\sigma_i| - \overline{\gamma}_i(\varepsilon_i)\right]. \tag{50}$$

Requiring $\dot{V} \leq -\sum_{i=1}^{3} \tilde{\mu}_i |\sigma_i|$ for eq. (50) the following domain $|\sigma_i| < \frac{\zeta_0}{\mu \lambda_{\min}} [\tilde{\mu}_i + \gamma_i(\varepsilon_i)] = \chi_i(\varepsilon_i)$ $\forall i = \overline{1,3}$ will be

reached in finite time $t_r = \max_{i=1,3} \frac{|\sigma_i(0)|}{\tilde{\mu}_i}$, where $\tilde{\mu}_i$ are chosen to provide given reaching times $t_{ri} = \frac{|\sigma_i(0)|}{\tilde{\mu}_i}$. Tracking

performance χ-robustness in the outer loop may be proved from the following expression

$$\left| e_i + c_i^1 \int_0^t e_i d\tau \right| \leq \chi_i(\varepsilon_i), \quad e_i = y_{ci} - y_i \;\; \forall i = \overline{1,3}, \tag{51}$$

in the sense that there exist[19] $\overline{g}_i > 0$ and $\overline{d}_i > 0$ that $|e_i| \leq \overline{g}_i |\sigma_i| + \overline{d}_i \leq \overline{g}_i \frac{\zeta_0}{\mu \lambda_{\min}} [\tilde{\mu}_i + \gamma_i(\varepsilon_i)] + \overline{d}_i$ within a sliding

surface domain of attraction ($|\sigma_i| \leq \chi_i(\varepsilon_i)$) in the outer loop. The value $\zeta_0 > 0$ must be chosen small enough and

the value $\mu > 0$ must be chosen large enough to achieve given tracking accuracy of mission angle command

profiles. The proposed algorithm selects the minimum boundary layer ε_i possible without violation of the actuator

limits. This will provide graceful tracking performance degradation as modeling error increases or control power

diminishes such as during damage or aggressive maneuvering.

Remark 5. Analyzing the system's steady state tracking performance in the inner loop it is easy to show that $\lim_{t\to\infty} \eta_i(t) = 0 \ \forall i = \overline{1,3}$ if $\Delta_{eq} = v_{eq} - \hat{v}_{eq} \to const$. Further, the system's steady state tracking error in the outer loop is as follows: $\lim_{t\to\infty} e_i(t) = 0 \ \forall i = \overline{1,3}$ if $\overline{\Phi}(.) \to const$ and $E(.) \to const$.

Examples

Consider the following nonlinear model of an F-16 jet fighter at $Mach = 0.7$, $h = 10000 \ ft$, $\alpha_{trim} = \theta_{trim} = 0.106803 rad, \delta_{e_{trim}} = -0.0295 rad$, $\beta_{trim} = p_{trim} = q_{trim} = r_{trim} = \varphi_{trim} = \delta_{a_{trim}} = \delta_{r_{trim}} = 0$

$$\dot{\theta} = q\cos\varphi - r\sin\varphi \tag{52}$$

$$\begin{cases} \dot{\varphi} = p + q\sin\varphi\tan\theta + r\cos\varphi\tan\theta \\ \dot{\alpha} = -\beta p + 0.0427\cos\theta\cos\varphi + 0.083589 + \tilde{Z}_\alpha\alpha + \tilde{Z}_q q + \tilde{Z}_\delta\delta_e \\ \dot{\beta} = -0.9973r + \alpha p + 0.0427\cos\theta\sin\varphi + \tilde{Y}_\beta\beta + \tilde{Y}_p p + \tilde{Y}_{\delta r}\delta_r + \tilde{Y}_{\delta a}\delta_a \end{cases} \tag{53}$$

$$\begin{cases} \dot{p} = -0.1345pq - 0.8225qr + \tilde{L}_\beta\beta + \tilde{L}_p p + \tilde{L}_r r - 50.933\delta_a + \tilde{L}_{\delta r}\delta_r \\ \dot{q} = 0.9586pr - 0.0833(r^2 - p^2) - 1.94166 + \tilde{M}_\alpha\alpha + \tilde{M}_q q + \tilde{M}_\delta\delta_e \\ \dot{r} = -0.7256pq + 0.1345qr + \tilde{N}_\beta\beta + \tilde{N}_p p + \tilde{N}_r r + 4.125\delta_a + \tilde{N}_{\delta r}\delta_r \end{cases} \tag{54}$$

$$\dot{\delta}_a = 20(u_2 - \delta_a), \quad \dot{\delta}_e = 20(u_1 - \delta_e), \quad \dot{\delta}_r = 20(u_3 - \delta_r), \tag{55}$$

Actuator deflection and rate limits are given by the following

$$|\delta_i| \le 0.37 rad, \quad |\dot{\delta}_i| \le 1 rad/s \quad \forall i = a, e, r. \tag{56}$$

Desired flying qualities[21] are defined by the following filters

$$\frac{\alpha_c}{\alpha^*} = \frac{4}{s^2 + 3s + 4}, \quad \frac{\beta_c}{\beta^*} = \frac{16.98}{s^2 + 6.18s + 16.98}, \quad \frac{\varphi_c}{\varphi^*} = \frac{2}{s^2 + 2.2s + 2}. \tag{57}$$

A 50% loss of horizontal tail and rudder areas is used to model aircraft battle damage. The parameter values given symbolically in eqs. (53) and (54) are represented as nominal terms $(.)_n$ and additive deviations from nominal due to the damage and uncertainty $\Delta(.)$[3,4]:

$$\tilde{G} = G_n + \Delta G, \tag{58}$$

where

$$G_n = \left[Z_\alpha, Z_q, Z_\delta, Y_\beta, Y_p, Y_{\delta r}, Y_{\delta a}, L_\beta, L_p, L_r, L_{\delta r}, M_\alpha, M_q, M_\delta, N_\beta, N_p, N_r, N_{\delta r}\right] = $$
$$[-1.15, 0.9937, 0.0, -0.297, 0.00085, 0.0, 0.0, -53.48, -4.324, -0.224,$$
$$10.177, 3.724, -1.26, -19.5, 17.67, 0.234 - 0.649, -6.155],$$

$$\Delta G = [\Delta Z_\alpha, \Delta Z_q, \Delta Z_\delta, \Delta Y_\beta, \Delta Y_p, \Delta Y_{\delta r}, \Delta L_\beta, \Delta L_p, \Delta L_r, \Delta L_{\delta r}, \Delta M_\alpha, \Delta M_q, \Delta M_\delta, \Delta N_\beta,$$
$$\Delta N_p, \Delta N_r, \Delta N_{\delta r}] = [0.04 \cdot 1(t-5), 0.0031 \cdot 1(t-5), -0.177 + 0.0885 \cdot 1(t-5),$$
$$0.0534 \cdot 1(t-1.5), -0.00005 \cdot 1(t-5), 0.0372 - 0.0186 \cdot 1(t-1.5), 0.002466, 8.024 \cdot 1(t-1.5),$$
$$0.071 \cdot 1(t-1.5), 0.055 \cdot 1(t-1.5), -5.089 \cdot 1(t-1.5), 1.856 \cdot 1(t-5), 0.42 \cdot 1(t-5),$$
$$9.75 \cdot 1(t-5), -5.82 \cdot 1(t-1.5), 0.01 \cdot 1(t-1.5), 0.133 \cdot 1(t-1.5), 3.077 \cdot 1(t-1.5)], \tag{59}$$
$$1(\tau) \text{ is a unit-step function.}$$

Two flight control problems are considered for this aircraft model. First, a simple longitudinal motion example is considered in which tracking of a pitch rate command $q_c(t)$ is desired before and after damage to the aircraft. The second example is a flight control problem for all axes in which de-coupled tracking of angle of attack, roll and sideslip angle commands is desired before and after damage to the aircraft.

Problem 1. The following simplified model of the aircraft longitudinal axis dynamics is derived[3,4] from the model given in eqs. (52)-(59):

$$\begin{bmatrix} \dot{\alpha} \\ \dot{q} \end{bmatrix} = \tilde{A}\begin{bmatrix} \alpha \\ q \end{bmatrix} + \tilde{B}\delta_e, \quad \dot{\delta}_e = -20\delta_e + 20u_1 , \tag{60}$$

where

$$\tilde{A} = A_n + \Delta A, \quad \tilde{B} = B_n + \Delta B ,$$

$$\begin{cases} A_n = \begin{bmatrix} Z_\alpha & Z_q \\ M_\alpha & M_q \end{bmatrix} = \begin{bmatrix} -1.1500 & 0.9937 \\ 3.7240 & -1.2600 \end{bmatrix}, \quad \Delta A = \begin{bmatrix} \Delta Z_\alpha & \Delta Z_q \\ \Delta M_\alpha & \Delta M_q \end{bmatrix} = \begin{bmatrix} 0.0400 & 0.0031 \\ 1.8560 & 0.4200 \end{bmatrix} \cdot 1(t-5), \\[4mm] B_n = \begin{bmatrix} Z_\delta \\ M_\delta \end{bmatrix} = \begin{bmatrix} 0 \\ -19.5000 \end{bmatrix}, \quad \Delta B = \begin{bmatrix} \Delta Z_\delta \\ \Delta M_\delta \end{bmatrix} = \begin{bmatrix} -0.177 + 0.0885 \cdot 1(t-5) \\ 9.7500 \cdot 1(t-5) \end{bmatrix}. \end{cases} \tag{61}$$

The elevator rate is limited by $\left| \dot{\delta}_e \right| \le b, \ (b = 1rad/s)$, and its deflection is limited by $\left| \delta_e \right| \le a, \ (a = 0.37rad)$. A "virtual" control is defined as

$$v_1 = -20\delta_e + 20u_1 , \tag{62}$$

and it is obvious that v_1 must obey the following inequality: $|v_1| \le b$.

A pitch rate command $q_c(t)$ is generated by a reference model of the form in eq. (60) that addresses desired flying qualities[21] where matrices A_r, B_r are the reference model state-space matrices[3,4]:

$A_r = \begin{bmatrix} -1.2693 & 0.9531 \\ -9.4176 & -5.7307 \end{bmatrix}, \ B_r = \begin{bmatrix} -0.1770 \\ -19.5000 \end{bmatrix}$. The pilot's commands u_{com} are 0.1 rad/s pitch rate pulses of 1.0 sec duration with polarities of -, +, -, + at times of 0.0, 3.0, 6.0, and 9.0 sec. respectively. A 50% loss of rudder area occurs at 5.0 sec.

A continuous SMC is designed with boundary layer re-configuration in the format of eq. (21) and eq. (36) without integral term to provide given settling time of a transient response in a sliding mode $t_s \le 1.0 s$. The coefficient c_2 is chosen $c_2 = 5 \ge \dfrac{4}{t_s}$, and the SMC is designed as follows:

$$u_1 = \delta_e + 0.05\rho_1 sat\frac{s_1}{\varepsilon_1}, \quad s_1 = \dot{\eta}_1 + 5.0\eta_1, \quad \eta_1 = q_c - q . \tag{63}$$

It is assumed that $\hat{v}_{1eq} = 0$. A value of the gain ρ_1 must be chosen to satisfy inequality (33) with

$$\rho_1 > \left| v_{1eq} \right| + \tilde{\rho}_1 , \quad \tilde{\rho}_1 > 0. \tag{64}$$

The equivalent control is assumed to have the following structure

$$v_{1eq} = v_{1eq_{nom}} + \tilde{v}_{1eq},$$ (65)

where $v_{1eq_{nom}}$ corresponds to the equivalent control in the nominal case, and $\tilde{v}_{1eq}$ is due to model uncertainty and aircraft damage. The value of $\tilde{\rho}_1$ in eq. (64) must be chosen to provide given reaching time $t_r \leq 1.0\,s$. This is[10-13]

$$\tilde{\rho}_1 \geq \frac{|s(0)|}{t_r}.$$ The value of $\tilde{\rho}_1 = 0.5$ is chosen for all $|s(0)| \leq 0.5$. The value of ρ_1 is chosen assuming that

$|\tilde{v}_{1eq}| < |v_{1eq_{nom}}|$. This is

$$\rho_1 > \left(|v_{1eq_{nom}}| + |\tilde{v}_{1eq}|\right) + 0.5 \Rightarrow \rho_1 = 3.0|v_{1eq_{nom}}| + 0.5,$$ (66)

$$v_{1eq_{nom}} = -0.05128(\ddot{q}_c + 5\dot{q}_c) + 0.495\alpha - 0.0519q - 3.73\delta_e.$$ (67)

Substituting $\delta_m = 0.37, \bar{\delta}_m = 1.0$ and $r^1 = r^2 = r^3 = 0.02$ into eqs. (45) and (46) gives the following expressions

$$\varepsilon_1^2 = \frac{0.05|s_1|\rho_1}{0.37 - \delta_e\, sign\, s_1} + 0.02, \qquad \varepsilon_1^3 = |s_1|\rho_1 + 0.02.$$ (68)

So, the re-configurable continuous SMC is completely defined by eqs. (40), (41), (63), (66)-(68). The simulation results are shown in figures 1-3. The $q_c(t)$ profile is tracked almost perfectly before and after a 50% loss of horizontal tail area (fig. 1). The actuator deflection and deflection rate are within the limits (fig.2). The tracking performance is much better when compared to the corresponding results from past work[3,4], which used explicit parameter identification for reconfiguration after damage. Further, the re-configurable continuous SMC is less complex since on-line identification is not required. The simulation was repeated with measurement noise. To compensate for the noise, a low-pass filter with a transfer function $W(s) = \dfrac{1}{0.03s + 1}$ filtered the sliding surface, and the gain ρ_1 was fixed at $\rho_1 = 3.0 \cdot \max|v_{1eq_{nom}}| + 0.5$. The tracking performance was similar to the case presented in fig. 1, and actuator deflection and deflection rate were within limits (fig. 4).

Problem 2. The multiple time scale concept is employed to design a two-loop control system with continuous SMC in both loops such that the attack, roll and sideslip angles follow commands before and after a 50% loss of horizontal tail and rudder.

A continuous SMC is first designed in the outer-loop. Nonlinear eqs. (52) and (53) of the F-16 mathematical flight model are considered. The angular rate command profiles p_c, q_c and r_c are treated as virtual controls for the command angles φ, α and β. The outer-loop SMC design starts with the choice of a sliding surface in eq. (6), given a settling time of a transient response in sliding mode $t_s \leq 1.5\,s$. The coefficients C_i^1 are chosen as follows: $C_i^1 = 3 \geq \dfrac{4}{t_s}$, and the sliding surface is designed

$$\sigma_i = e_i + 3.0\int_o^t e_i d\tau, \quad \forall i = \overline{1,3},$$ (69)

where $e_1 = \varphi_c - \varphi$, $e_2 = \alpha_c - \alpha$, $e_3 = \beta_c - \beta$. A SMC is chosen in the form of eq. (11) with $\hat{z}_{c_{eq}} = \{\hat{p}_{eq}, \hat{q}_{eq}, \hat{r}_{eq}\}^T$ calculated as follows:

$$\begin{cases} \hat{p}_{eq} = \dot{\varphi}_c + 3.0\varphi_c - q\sin\varphi\tan\theta - r\cos\varphi\tan\theta - 3.0\varphi \\ \hat{q}_{eq} = \dot{\alpha}_c + 3\alpha_c - 1.85\alpha + \beta p - 0.0427\cos\theta\cos\varphi - 0.083589 \\ \hat{r}_{eq} = -\dot{\beta}_c - 3.0\beta_c + 2.7029\beta + \alpha p + 0.0427\cos\theta\sin\varphi + 0.000851p. \end{cases}$$

The following matrices are constructed

$$B_1(.) = \begin{bmatrix} 1 & \sin\varphi\tan\theta & \cos\varphi\tan\theta \\ -\beta & 0.9937 & 0 \\ \alpha + 0.00085 & 0 & -0.9973 \end{bmatrix}, \quad \Delta B_1(.) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.0031 & 0 \\ -0.00005 & 0 & 0 \end{bmatrix} \text{ and}$$

$$E = \Delta B_1 \cdot B_1^{-1} = \frac{10^{-3}}{|B_1|} \begin{bmatrix} 0 & 0 & 0 \\ -3.1\beta & -0.9973 - (\alpha + 0.00085)\cos\varphi\tan\theta & -\beta\cos\varphi\tan\theta \\ 4.97\cdot10^{-2} & -4.99\cdot10^{-2}\sin\varphi\tan\theta & 4.97\cdot10^{-2}\cos\varphi\tan\theta \end{bmatrix}$$

where $|B_1| = -0.991 - 0.9937(\alpha + 0.00085)\cos\varphi\tan\theta - 0.9973\sin\varphi\tan\theta$. The elements of the E matrix satisfy the conditions of eq. (15) in a significant flight envelope. For relatively small angles β, φ, α and θ, the matrix B_1 is diagonal-dominant. This allows outer-loop SMC (eqs. (11) and (18)) design in a simplified de-coupled format. Finally, the outer-loop SMC is designed in the following continuous format:

$$p_c = \hat{p}_{eq} + 15.0\sigma_1, \quad q_c = \hat{q}_{eq} + 50.0\sigma_2, \quad r_c = \hat{r}_{eq} - 6.0\sigma_3. \tag{70}$$

Given the reference profiles p_c, q_c, r_c in eq. (70) a re-configurable continuous SMC is now designed for the inner loop. Recall that $\tilde{B}_2$ must be positive definite for damage and undamaged cases. The matrix

$$-B_2 = \begin{bmatrix} 50.933 & 0 & -10.177 \\ 0 & 19.5 & 0 \\ -4.125 & 0 & 6.155 \end{bmatrix} \text{ is diagonal-dominant and positive definite in accordance with Gershgorin}$$

theorem[19]. Note that a control law command sign change justifies analyzing the positive definiteness of $-B_2$ instead of B_2 ($B_2 u = (-B_2)(-u)$). The damaged matrix $-\tilde{B}_2 = \begin{bmatrix} 50.933 & 0 & -5.089 \\ 0 & 9.75 & 0 \\ -4.125 & 0 & 3.077 \end{bmatrix}$ is not diagonal dominant,

however it is still positive definite in accordance with Sylvester criterion[13,19].

The sliding surfaces are designed in eq. (21) format by considering the F-16 mathematical flight model in eqs. (54) and (55) and given settling time of a transient response in the inner loop sliding mode $\tilde{t}_s \leq 0.75s$ that is twice as fast as in the outer loop. This is

$$s_1 = \dot{\eta}_1 + 12\eta_1 + 36\int_0^t \eta_1 d\tau, \quad s_2 = \dot{\eta}_2 + 12\eta_2 + 36\int_0^t \eta_2 d\tau, \quad s_3 = \dot{\eta}_3 + 16\eta_3 + 64\int_0^t \eta_3 d\tau \tag{71}$$

where $\eta_1 = p_c - p$, $\eta_2 = q_c - q$, $\eta_3 = r_c - r$.

The control functions are designed in the format of eq. (36) assuming $\hat{v}_{eq_i} = 0$

$$u_i = \delta_i + 0.05\rho_i sat\frac{s_i}{\varepsilon_i}, \quad \rho_i > \left|v_{eq_i}\right| \quad \forall i = \overline{1,3} \ . \tag{72}$$

Addressing the actuator deflection and deflection rate limits in eq. (56), the boundary layer is reconfigured according to eqs. (40), (41, (45) and (46) as follows:

$$\varepsilon_i = \max\left\{\varepsilon_i^1, \varepsilon_i^2, \varepsilon_i^3\right\} \quad \forall i = \overline{1,3} \ , \tag{73}$$

$$\varepsilon_i^1 = \left|s_i\right| + 0.02, \quad \varepsilon_i^2 = \frac{0.05\left|s_i\right|\rho_i}{(0.37 - \delta_i signs_i)} + 0.02, \quad \varepsilon_i^3 = \left|s_i\right|\rho_i + 0.02, \quad \forall i = \overline{1,3} \ , \tag{74}$$

The following simulation commands are considered for evaluation of the two-loop re-configurable SMC:

- The angle of attack command (with respect to a trim value $\alpha_{trim} = 0.106803 \, rad$) is two 3.0 s duration pulses of 0.1 rad amplitude and polarities + and - during time intervals [0,3] and [6, 9] seconds respectively.

- The roll angle command is one pulse of 1.0 rad amplitude during the time interval of [1, 9] seconds.

- The sideslip angle command is zero for all time.

The flying quality filters in eq. (57) are used to filter the commands. Loss of 50% of rudder area occurs at 1.5 sec, and loss of 50% of tail area occurs at 5.0 sec. The results of the simulation given in figures 5-9 show that the prescribed maneuver is accomplished with excellent tracking performance before and after damage to the aircraft without violating the actuator limits. Figure 10 shows that the boundary layer thickness varies significantly during aggressive commands and once aircraft damage occurs yet remains bounded.

Conclusions

A re-configurable two-loop continuous SMC method is developed and successfully applied to a realistic aircraft flight control problem. Actuator limit compliance is achieved by sliding surface boundary layer direct adaptation and demonstrated through high fidelity simulations of an aircraft subject to major damage and measurement noise.

II.   Finite-Reaching-Time Continuous Sliding Mode Controller for MIMO Nonlinear Systems

Introduction

Sliding mode control is a nonlinear robust control technique, which is actively developing during last 30-35 years[10,11]. A sliding mode controller design comprises two steps[10,11]. At the first step a "custom-made" surface (manifold) is to be designed in a state variable space such that the closed-loop system exhibits a given (desired) performance while moving (sliding) on this surface, called sliding surface. While on the sliding surface the plant's dynamics is restricted to the equations of the surface and are robust to matched plant uncertainties and external disturbances. At the second step a feedback control law is to be designed to provide convergence of a system's trajectory to the sliding surface. In particularly the designed control law must provide an asymptotic stability of the origin in a sliding surface subspace. Since the closed-loop system exhibits he desired robust performance moving on

the sliding surface, it is highly desirable to reach the sliding surface in a finite time. The system's motion on the sliding surface is called the sliding mode. Historically, sliding modes were discovered in variable structure systems, which feedback structure changes depending on value of a state vector to drive the system's trajectory onto the "custom-made" sliding surface. While on the surface the variable structure control yields an infinity frequency switching control[10,11], which is practically unrealizable. Dynamics of switching elements and finite sampling time of microprocessors prevent from implementation of infinity frequency switching control law and lead to so called control chattering (limit cycles with finite frequency oscillation in a closed-loop system). In order to avoid chattering a continuous *implementation* of a discontinuous control law is often considered[12-14,22]. Usually this is an equivalent control plus a linear control law with respect to sliding surface. Very often a linear control law is taken with saturation[13,14]. The continuous implementations of a sliding mode controller (SMC) are widely known for instance[16,22-26]. Such SMCs are free from chattering, but have some disadvantages comparing with infinity frequency switching sliding mode controllers. They are:

1. In undisturbed case a sliding surface can be reached only asymptotically, i.e. reaching time $\rightarrow \infty$.

2. In case of nonvanishing disturbances and uncertainties convergence is provided only to some vicinity of the sliding surface. It means loss of robustness comparing with infinity frequency switching sliding mode controller.

In this work we consider a *continuous sliding mode controller* as a specific type of a SMC[22,27,28], but not as implementation of a discontinuous high frequency switching control law. Continuous SMCs, which drive trajectories of a single input-single-output (SISO) and MIMO nonlinear control system to sliding surfaces in a finite time are designed in the works[15,29]. In this report we follow the work[29], in which a finite-reaching-time continuous sliding mode controller is designed for multi-input-multi-output (MIMO) nonlinear control systems using finite-time differential equations analysis[30,31]. Robustness of the designed controller, which is an issue of crucial importance, is addressed as well. Theoretical results are validated via computer simulations.

Problem Formulation

Consider the nonlinear square MIMO system[13]

$$\dot{x} = f(x) + G(x)u, \quad y = h(x) \tag{75}$$

where $x \in R^n$, $y \in R^m$, $u \in R^m$. Assume that the functions $f(x)$, $h(x)$ are smooth vector fields, and columns $g_i(x)$ $\forall i = \overline{1,m}$ of the matrix $G(x) \in R^{n \times m}$ are smooth vector fields as well. Assume also that the system (75) is completely linearizable in a reasonable domain $x \in \Gamma$, i.e. the system (75) does not have internal (zero) dynamics.

Given in real time an output reference profile $y_r(t)$ identify a continuous sliding mode controller which

1. provides a given eigenvalues to the de-coupled output tracking error performance in the sliding mode,

2. drives the closed-loop system trajectory to the sliding surface in a finite time.

System Transformation

Differentiating the outputs, the system (75) can be easily transformed to a normal form[13]

$$\begin{bmatrix} y_1^{(r_1)} \\ y_2^{(r_2)} \\ \text{.......} \\ y_m^{(r_m)} \end{bmatrix} = \begin{bmatrix} L_f^{r_1} h_1(x) \\ L_f^{r_2} h_2(x) \\ \text{............} \\ L_f^{r_m} h_m(x) \end{bmatrix} + E(x)u, \qquad E(x) = \begin{bmatrix} L_{g_1}(L_f^{r_1-1}h_1) \; L_{g_2}(L_f^{r_1-1}h_1) \ldots L_{g_m}(L_f^{r_1-1}h_1) \\ L_{g_1}(L_f^{r_2-1}h_2) \; L_{g_2}(L_f^{r_2-1}h_2) \ldots L_{g_m}(L_f^{r_2-1}h_2) \\ \text{...} \\ L_{g_1}(L_f^{r_m-1}h_2) \; L_{g_2}(L_f^{r_m-1}h_m) \ldots L_{g_m}(L_f^{r_m-1}h_m) \end{bmatrix} \qquad (76)$$

where $|E(x)| \neq 0 \; \forall x \in \Gamma$; $L_f^{r_i} h_i$ and $L_{g_i}(L_f^{r_i-1}h_i) \; \forall i = \overline{1,m}$ are corresponding Lie derivatives.

Terminal Attractor

Following the works[29-31], a useful result is formulated in a following Lemma.

Lemma. The origin of a differential equation

$$\dot{z} = -\omega z^a, \quad z(0) > 0, \omega > 0 \qquad (77)$$

is a terminal attractor with a finite reaching time

$$t_r = \frac{(z(0))^{1-a}}{\omega(1-a)} \quad \forall a \in (0,1). \qquad (78)$$

Indeed, integrating (3) we have obtained

$$(z(t))^{1-a} = (z(0))^{1-a} - (1-a)\omega t. \qquad (79)$$

Requiring $z(t_r)=0$ implies formula (78).

Finite-Reaching-Time SMC Design for Nonlinear MIMO Output Tracking

Achieving de-coupled output tracking error performance in the sliding mode, the following sliding surfaces are designed

$$\sigma_i = e_i^{(r_i-1)} + c_{i,r_i-2}e_i^{(r_i-2)} + \ldots + c_{i,1}e_i^{(1)} + c_{i,0}e_i, \; \forall i = \overline{1,m}, \qquad (80)$$

where $e_i = y_{r,i}(t) - y_i(t)$, $e_i^{(j)} = \dfrac{d^j e_i}{dt^j}$ and the coefficients $c_{i,j} \; \forall i = \overline{1,m} \forall j = \overline{0, r_i - 2}$ are chosen to provide given

eigenvalues placement to the de-coupled differential equations $\sigma_i = 0 \; \forall i = \overline{1,m}$.

Designing a continuous sliding mode controller for the system (76) we need to provide asymptotic stability to its motion in the σ-subspace, which is described as follows:

$$\dot{\sigma} = \Psi(.) - E(x)u \qquad (81)$$

where

$$\Psi(.) = \{\psi_1(.), \psi_2(.), \ldots, \psi_m(.)\}^T, \; \psi_i(.) = y_{r,i}^{(r_i)}(t) + c_{i,r_i-2}e_i^{(r_i-1)} + \ldots + c_{i,0}e_i^{(1)} - L_f^{r_i}h_i(x) \; \forall i = \overline{1,m}. \qquad (82)$$

Introducing a new control variable

$$\tilde{u} = E(x)u, \qquad (83)$$

the system (81) can be rewritten in a scalar format. This is

$$\dot{\sigma}_i = \psi_i(.) - \tilde{u}_i \quad \forall i = \overline{1,m} \qquad (84)$$

The system (84) is obviously input-output de-coupled. Hence, analyzing stability of the origin of the system (84), we can analyze stability of each ith equation $\forall i = \overline{1,m}$. Candidates to Lyapunov functions are introduced

$$V_i = \frac{1}{2}\sigma_i^2 \tag{85}$$

We will look for the derivative of the candidates to Lyapunov functions (85) in the following format:

$$\dot{V}_i = -\omega_i V_i^{a_i}, \quad a_i > 0, \, \omega_i > 0. \tag{86}$$

Then, applying Lemma we can conclude that the origin of the ith equation of the system (84) will be reached in a finite time $\forall i = \overline{1,m}$

$$t_{r_i} = \frac{(V_i(0))^{1-a_i}}{\omega_i(1-a_i)} \quad \forall a_i \in (0,1) \tag{87}$$

It means that $\sigma_i = 0$ $\forall i = \overline{1,m}$ become terminal attractors. Substituting expressions (80) and (84) into formulas (85), a control function $\tilde{u}$ is identified as follows:

$$\tilde{u} = -\Psi(.) + R \cdot \Sigma, \tag{88}$$

where $R = diag\left\{\dfrac{\omega_i}{2^{a_i}}\right\}$ $\forall i = \overline{1,m}$; $\Sigma = \left\{\sigma_1^{2a_1-1}, \sigma_2^{2a_2-1}, ..., \sigma_m^{2a_m-1}\right\}^T$.

Substituting equation (88) into equations (84), we obtained equations of the closed-loop system's (76) motion in the σ-subspace. They are

$$\dot{\sigma}_i = -2^{-a_i}\omega_i\sigma^{2a_i-1} \quad \forall i = \overline{1,m}. \tag{89}$$

Making coordinates of the control function (88) continuous and avoiding singularity in the origin of the system (89), we will look for the parameters "a_i" in the following format:

$$a_i = \frac{p_i}{q_i} \quad \forall i = \overline{1,m}, \tag{90}$$

where p_i and q_i are integer numbers. Since σ_i can be either positive or negative the values of q_i and $2p_i - q_i > 0$ must be odd. Hence, the parameters p_i and q_i must meet the following conditions

$$p_i, q_i \in \Omega_i: \left\{p_i, q_i \middle| 2p_i > q_i, \, p_i < q_i, \, q_i \text{ is odd}\right\} \forall i = \overline{1,m}. \tag{91}$$

As soon as the values of the parameters p_i and q_i have been identified in accordance with the conditions (17) the value of the parameters ω_i must be chosen to achieve a given reaching times (87).

The original finite-reaching-time continuous sliding mode control law u has been identified in accordance with formulas (83) and (88) as follows:

$$u = E(x)^{-1}\left[-\Psi(.) + R \cdot \Sigma\right] \tag{92}$$

The obtained results are formulated into the following Theorem.

Theorem. The sliding surfaces (84) will be reached in finite times (87) via the continuous sliding mode controller (92) if values of parameters "a_i" are chosen in accordance with formulas (90) and (91).

Robustness

Assuming

$$\Psi(.) = \Psi_0(.) + \Delta\Psi(.), \quad E(x) = E_0(x) + \Delta E(x) \tag{93}$$

where $\Delta\Psi(.)$ and $\Delta E(x)$ are uncertain vector and matrix correspondingly. Designing control (92) in the following format

$$u = E_0(x)^{-1}\left[-\Psi_0(.) + R\cdot\Sigma\right], \tag{94}$$

and substituting expression (94) into equations (84), we obtained

$$\dot{\sigma} = \Delta\Psi(.) - \Delta M \cdot \Psi_0(.) - (I + \Delta M)R\cdot\Sigma, \tag{95}$$

where $\Delta M = \Delta E(x)\cdot E_0^{-1}(x)$.

Equation (95) is rewritten in a scalar format. This is

$$\dot{\sigma}_i = \Delta\psi_i(.) - (\Delta M\cdot\Psi_0(.))_i - \frac{\omega_i}{2^{a_i}}\sigma_i^{2a_i-1} - \sum_{j=1}^{m}\Delta M_{ij}\frac{\omega_j}{2^{\omega_j}}\sigma_j^{2a_j-1} \quad , \forall i = \overline{1,m} \tag{96}$$

Substituting equations (96) into equations (85), we obtained

$$\dot{V}_i = -\left[\omega_i(1+\Delta M_{ii}) + \frac{2^{a_i}\left(-\Delta\psi_i(.) + (\Delta M\cdot\Psi_0(.))_i + \sum_{\substack{j=1\\j\neq i}}^{m}\Delta M_{ij}\frac{\omega_j}{2^{\omega_j}}\sigma_j^{2a_j-1}\right)}{\sigma_i^{2a_i-1}}\right]V_i^{a_i}. \tag{97}$$

Assuming

$$|\Delta M_{ii}| \le L_{ii} < 1, \quad |\Delta M_{ij}| \le L_{ij}, \quad |\Delta\psi_i| \le R_i, \quad |(\Delta M\cdot\Psi_0(.))_i| \le G_i, \quad |\sigma_j| \le \pi_j \quad \forall i = \overline{1,m} \,\, \forall j = \overline{1,m} \tag{98}$$

we introduce positive constants

$$\Phi_i = R_i + G_i + \sum_{\substack{j=1\\j\neq i}}^{m}\frac{\omega_j}{2^{a_j}}L_{ij}\pi_j^{2a_j-1} \tag{99}$$

Transforming equations (97) taking into account expressions (98) and (99) we obtained

$$\dot{V}_i \le -\left[\omega_i(1-L_{ii}) - \frac{2^{a_i}\Phi_i}{\left|\sigma_i^{2a_i-1}\right|}\right]V_i^{a_i} \quad \forall i = \overline{1,m}. \tag{100}$$

Inequalities (100) are rewritten in a format

$$\dot{V}_i \le -\tilde{\omega}_i V_i^{a_i} \quad \forall i = \overline{1,m}, \tag{101}$$

where

$$\tilde{\omega}_i = \frac{(V_i(0))^{1-a_i}}{t_{r_i}(1-a_i)}, \forall i = \overline{1,m} \tag{102}$$

and

$$\omega_i(1 - L_{ii}) - \frac{2^{a_i}\Phi_i}{\left|\sigma_i^{2a_i-1}\right|} \geq \tilde{\omega}_i \quad \forall i = \overline{1,m}. \tag{103}$$

Inequalities (103) actually form domains where conditions (101) are met. It means that the following domains

$$|\sigma_i| \leq \left(\frac{2^{a_i}\Phi_i}{\omega_i(1 - L_{ii}) - \tilde{\omega}_i}\right)^{\frac{1}{2a_i-1}} \quad \forall i = \overline{1,m} \tag{104}$$

will be reached by the trajectories σ_i in given finite times (87), and trajectories σ_i will stay in these domains thereafter. Hence, solutions of the closed loop systems (96) are uniformly ultimately bounded[14]. What is remarkable, that choosing values of the parameters ω_i to meet inequalities

$$\omega_i > \frac{2^{a_i}\Phi_i + \tilde{\omega}_i}{1 - L_{ii}} \quad \forall i = \overline{1,m} \tag{105}$$

and making $2a_i - 1$ arbitrary small, we can make a size of the domains of attraction in eq. (104) arbitrary small as well, retaining the control law (94) continuous. In order to achieve a similar result, using linear/saturation approximation of a discontinuous sliding mode controller, a gain of such controller must be arbitrary large[13,14], which is impractical. Hence, a designed continuous sliding mode controller is extremely robust.

Example

A nonlinear MIMO system given in the format (75) is considered

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = \cos x_3 + \left(x_1^2 + 1\right)u_1 + u_2 + f_1(t) \\ \dot{x}_3 = \sin x_1 + 0.5u_1 + \left(x_2^2 + 1\right)u_2 + f_2(t) \\ y_1 = x_1, \quad y_2 = x_3 \end{cases} \tag{106}$$

Functions $f_i(t) : |f_i(t)| \leq \eta_i \quad \forall i = 1,2$ are unknown, bounded, matched disturbances. The system (106) is completely linearizable with a vector-relative degree $r = \{2, 1\}$. The system (106) is transformed[12,13] into the normal format (76) as follows:

$$\begin{bmatrix} \ddot{y}_1 \\ \dot{y}_2 \end{bmatrix} = \begin{bmatrix} \cos x_3 \\ \sin x_1 \end{bmatrix} + \begin{bmatrix} f_1(t) \\ f_2(t) \end{bmatrix} + \begin{bmatrix} \left(x_1^2 + 1\right) & 1 \\ 0.5 & \left(x_2^2 + 1\right) \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}. \tag{107}$$

The goal is to design a robust, continuous sliding mode controller to provide de-coupled tracking of smooth output reference profiles $y_{1r}(t)$ and $y_{2r}(t)$ in finite-reaching-time sliding modes.

The sliding surfaces are designed in the format (80) as follows

$$\sigma_1 = \dot{e}_1 + 5e_1, \quad \sigma_2 = e_2, \tag{108}$$

where $e_1 = y_{1r} - y_1$, $e_2 = y_{2r} - y_2$.

The parameters p_i, q_i and ω_i are identified $\forall i = 1,2$. They are

$$p_1 = p_2 = 2, \quad q_1 = q_2 = 3, \quad \omega_1 = 16, \quad \omega_2 = 10.$$

The continuous "cubic root" sliding mode controller is designed in the format (92) as follows:

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} \dfrac{x_2^2+1}{(x_1^2+1)(x_2^2+1)-0.5} & -\dfrac{1}{(x_1^2+1)(x_2^2+1)-0.5} \\ -\dfrac{0.5}{(x_1^2+1)(x_2^2+1)-0.5} & \dfrac{x_1^2+1}{(x_1^2+1)(x_2^2+1)-0.5} \end{bmatrix} \begin{bmatrix} \tilde{u}_1 \\ \tilde{u}_2 \end{bmatrix} \tag{109}$$

where

$$\begin{cases} \tilde{u}_1 = \ddot{y}_{1r} + 5\dot{y}_{1r} - \cos x_3 - 5x_2 + 10.0\sqrt[3]{\sigma_1} \\ \tilde{u}_2 = \dot{y}_{2r} - \sin x_1 + 6.3\sqrt[3]{\sigma_2} \end{cases} \tag{110}$$

The system (106) was simulated with the continuous "cubic root" sliding mode controller (110) and with the continuous "linear" sliding mode controller (111)

$$\begin{cases} \tilde{u}_1 = \ddot{y}_{1r} + 5\dot{y}_{1r} - \cos x_3 - 5x_2 + 16\sigma_1, \\ \tilde{u}_2 = \dot{y}_{2r} - \sin x_1 + 20\sigma_2. \end{cases} \tag{111}$$

It is worth to note that gains before the σ – terms in formulas (111) are increased from 10 to 16 and from 6.3 to 20 comparing with formulas (110), which helps the "linear" controller (111) to be more aggressive.

The following initial conditions and the reference profiles were used

$$x_1(0) = 0.2,\ x_2(0) = 0.1,\ x_3(0) = -0.15,\ y_{1r}(t) = 0.5\sin t,\ y_{2r}(t) = 0.5\cos t\ . \tag{112}$$

The results of the simulation are shown in fig. 11-17. Simulations confirm a finite-reaching time property of the designed continuous sliding mode controller and show improving robustness to disturbances comparing with "linear" sliding mode controller.

Conclusion

The continuous sliding mode controller, which reaches sliding surfaces in a finite time, is developed for nonlinear MIMO output tracking. Robustness of the designed controller is analytically analyzed and validated via computer simulations.

III.   Flight Control Using Re-configurable Sliding Modes for Tailess Aircraft

Introduction

Tailess aircraft has a reduced vertical tail or no vertical tail at all. Radar cross section and weight reduction benefits have influenced this tailess aircraft configuration. On the other hand, tailess configuration present a challenge from a stability and control perspective[32,33]. Absence of a vertical tail reduces directional stability and directional control power. Unconventional control effectors are usually added to overcome control power deficiencies[32]. The unconventional control effectors must be allocated with other conventional effectors to achieve acceptable maneuvering performance of a tailess fighter via a corresponding control law design.

A modular (multiple-loop) flight control structure for flight control of a tailess aircraft is developed[33,34]. It includes a feedback control, parameter estimation and control allocation modules. The control allocation module allocates the control effectors to achieve maximum control power and limit control law commands to prevent actuator saturation[32]. The developed[32] control allocation algorithm operates upon assumption that full information

about an aircraft mathematical model is available. A parameter estimation module supports this assumption. However, in case of a battle damage to an aircraft, parameter estimation becomes a "bottle-neck" problem, because errors in damage estimation can lead to a significant degradation of an aircraft performance.

In this work we develop a re-configurable flight control design technique on the basis of Sliding Mode Control for tailless aircraft with damages and actuator limits. On-line damage (parameter) estimation is not required. An optimal control allocation algorithm is employed using nominal mathematical model of an aircraft. Continuous SMC with direct adaptation of a boundary layer automatically adjusts to an aircraft damage and prevents actuator from deflection and rate saturation.

Problem Formulation

Consider a mathematical flight dynamics model of a rigid tailless aircraft with the following cascade structure

$$
\begin{cases}
\dot{y} = f_1(y,\xi) + \Delta f_1(y,\delta,\xi) + (B_1(y,\xi) + \Delta B_1(y,\xi))z \\
\dot{z} = f_2(y,z,\xi) + \Delta f_2(y,z,\xi) + (B_2(y,z,\xi) + \Delta B_2(y,z,\xi))\delta \\
\dot{\xi} = f_3(y,z,\xi,\delta) \\
\dot{\delta} = -A_\delta(\delta - u),
\end{cases} \tag{113}
$$

where $y \in \Re^3$ is a vector of command angles ($y = \{\varphi, \alpha, \beta\}^T$, and φ, α, β, are roll angle, angle of attack, angle of sideslip respectively); $z \in \Re^3$ is a vector of angular rates ($z = \{p,q,r\}^T$, and p,q,r are roll, pitch and yaw rates respectively); $\delta \in \Re^m$ is a vector of control surface deflections ($m \geq 3$), $u \in \Re^m$ is a control vector containing actuator inputs, $\xi \in \Re^k$ is a vector containing all other states not contained in y, z, or δ; $f_1(y,\xi) \in \Re^3$, $f_2(y,z,\xi) \in \Re^3$, $f_3(y,z,\xi,\delta) \in \Re^k$ are differentiable vector-functions; $B_1(y,\xi) \in \Re^{3\times3}$, $B_2(y,z,\xi) \in \Re^{3\times m}$ are known matrices; $\Delta f_1 \in \Re^3$, $\Delta f_2 \in \Re^3$, $\Delta B_1 \in \Re^{3\times3}$, $\Delta B_2 \in \Re^{3\times m}$ are unknown bounded smooth perturbations, caused by damage, failures and modeling uncertainties; $\det(B_1(y,\xi) + \Delta B_1(y,\xi)) \neq 0 \ \forall y,\xi \in \Omega$.

The deflections and deflection rates of the actuators are assumed to be bounded:

$$
|\delta_i| \leq \delta_m, |\dot{\delta}_i| \leq \bar{\delta}_m \ \forall i = \overline{1,m}. \tag{114}
$$

Problem: Given a real-time command reference profile $y_c(t)$, design a continuous SMC to obtain the robust asymptotic de-coupled motion of the output tracking errors

$$
\lim_{t \to \infty} |y_{ci} - y_i| = 0 \ \forall i = \overline{1,3} \tag{115}
$$

in sliding modes described by homogeneous time-invariant equations with given eigenvalues placement for the system in eqs. (113) and (114).

Assumption. It is assumed that the equations

$$
\dot{\xi} = f_3(y_c, z_c, \xi, \delta_c) \tag{116}
$$

are stable, where δ_c is the commanded deflection that satisfies equation $\dot{\delta}_c = -A_\delta(\delta_c - \tilde{u})$, and $\tilde{u}$ is the control that maintains commanded states y_c, z_c.

This is a reasonable assumption since if any elements of ξ were not stable, it is argued that those elements should be moved to either y or z depending on the dynamical structure. This assumption is analogous to the minimum phase assumption used in feedback linearization.

The formulated control problem is similar to the problem that was formulated and addresses in the first chapter of this report. The main difference is in non-square configuration of a tailess aircraft model in eq. (113).

Outer Loop Finite-Reaching Time Continuous SMC Design

Considering the angular rate vector z as a virtual control command z_c, only the command angle dynamics are considered for the outer loop SMC design

$$\dot{y} = f_1(y,\xi) + \Delta f_1(y,\delta,\xi) + (B_1(y,\xi) + \Delta B_1(y,\xi))z .$$
(117)

The system eq. (117) is already in a eq (76) format with the vector relative degree of eq. (117) is $\bar{r} = [1,1,1]^T$, and the sliding surface can be designed as follows[10-13]:

$$\sigma = e + C^1 \int_0^t e d\tau = 0, \qquad \sigma, e \in R^3 ,$$
(118)

where $e = \{e_1, e_2, e_3\}^T$, $e_i = y_{ci} - y_i$ is the command angle error and $C^1 = diag\{c_i^1\}$ $\forall i = \overline{1,3}$ is chosen to achieve desired command angle error dynamics. In order to have desired command angle error dynamics and to nullify a steady state tracking error due to continuous SMC the sliding surface is designed with an integral term in the PI format eq. (118).

The second step of the outer loop SMC design is to choose a control function that achieves asymptotic stability with finite reaching time of the σ dynamics given as

$$\dot{\sigma} = \Psi_0(.) + \Delta\Psi(.) - [B_1(.) + \Delta B_1(.)]z_c ,$$
(119)

where

$$\Psi_0(.) = \dot{y}_c - f_1(.) + C^1 e, \quad \Delta\Psi(.) = -\Delta f_1(.) .$$
(120)

Applying the finite-reaching time continuous SMC design technique that is developed in the Chapter II and using eq. (94), we obtained expression for the virtual control z_c as follows:

$$z_c = \hat{z}_{c_{eq}} + B_1^{-1}(.) \cdot R \cdot \Sigma ,$$
(121)

where

$$\hat{z}_{c_{eq}} = B_1^{-1}(.)\Psi_0(.), \quad R = diag\left\{\frac{\omega_1}{2^{a_1}}, \frac{\omega_2}{2^{a_2}}, \frac{\omega_3}{2^{a_3}}\right\}; \quad \Sigma = \{\sigma_1^{2a_1-1}, \sigma_2^{2a_2-1}, \sigma_3^{2a_3-1}\}^T .$$
(122)

Denoting $\Delta M(.) = \Delta B_1(.) \cdot B_1^{-1}(.)$, $g(.) = \Delta M(.) \cdot \Psi_0(.)$ and assuming

$$|\Delta M_{ii}| \leq L_{ii} < 1, \quad |\Delta M_{ij}| \leq L_{ij}, \quad |\Delta\Psi_i| \leq \lambda_i, \quad |g(.)_i| \leq G_i, \quad |\sigma_j| \leq \pi_j \quad \forall i = \overline{1,3} \ \forall j = \overline{1,3},$$
(123)

we uniform ultimate boundadness of σ-dynamics. In other words, the control law in eq. (121) guarantees convergence of σ − trajectories to a domain

$$|\sigma_i| \le \left(\frac{2^{a_i} \Phi_i}{\omega_i (1 - L_{ii}) - \tilde{\omega}_i} \right)^{\frac{1}{2a_i - 1}} \quad \forall i = \overline{1,3}, \tag{124}$$

where

$$\Phi_i = \lambda_i + G_i + \sum_{\substack{j=1 \\ j \ne i}}^{3} \frac{\omega_j}{2^{a_j}} L_{ij} \pi_j^{2a_j - 1} \quad \forall i = \overline{1,3}. \tag{125}$$

Parameters ω_i and a_i of the continuous finite-reaching-time SMC eq. (121) must be chosen to provide given reaching time in eq (102) and given size of the domain of convergence of σ – trajectories in eq. (124).

Inner Loop Finite-Reaching-Time Continuous SMC Design and Control Allocation.

The angular rate z – dynamics are considered for the inner loop SMC design

$$\dot{z} = f_2(y,z,\xi) + \Delta f_2(y,z,\xi) + (B_2(y,\xi) + \Delta B_2(y,\xi))\delta. \tag{126}$$

Vector-deflection δ is treated as a control function in eq. (126).

A continuous SMC δ_c should be designed to obtain the robust asymptotic de-coupled motion of the tracking errors

$$\lim_{t \to \infty} |z_{ci} - z_i| = 0 \quad \forall i = \overline{1,3} \tag{127}$$

in sliding modes described by homogeneous time-invariant equations with given eigenvalues placement for the system in eqs. (126) and (2). Note that the angular rate command z_c is generated by the outer loop SMC.

The vector deflection-command δ_c will be designed into two steps.

<u>At the first step</u>, denoting

$$B_2(y,\xi)\delta_c = v_c, \quad v_c \in \Re^3, \tag{128}$$

and rewriting eq. (126) in a following square structure:

$$\dot{z} = f_2(y,z,\xi) + \Delta f_2(y,z,\xi) + v_c + \Delta B_2(y,\xi)\delta_c, \tag{129}$$

we will design v_c in eq. (129) as a new control function to address a problem in eq. (127).

<u>At the second step</u>, assuming the control law v_c is designed in a finite-reaching-time continuous SMC format, a control allocation must be performed to transform the designed control law v_c into deflection vector-command δ_c:

$$\delta_c = B_v(.)v_c, \tag{130}$$

where $B_v \in \Re^{m \times 3}$ is an allocation matrix that must be identified.

Substituting eq. (130) into eq. (129) we obtained

$$\dot{z} = f_2(y,z,\xi) + \Delta f_2(y,z,\xi) + (I + E(.))v_c, \tag{131}$$

where $E(.) = \Delta B_2(.)B_v(.)$. Obviously, structure of eq. (131) is absolutely the same as in eq. (117). Therefore, the sliding surface is designed similar to eq. (118)

$$\bar{\sigma} = \bar{e} + \overline{C}^1 \int_0^t e \, d\tau = 0, \qquad \bar{\sigma}, \bar{e} \in R^3, \tag{132}$$

where $\bar{e} = \{\bar{e}_1, \bar{e}_2, \bar{e}_3\}^T$, $\bar{e}_i = z_{ci} - z_i$ is the command angle error and $\overline{C}^1 = diag\{\bar{c}_i^1\}$ $\forall i = \overline{1,3}$ is chosen to achieve desired angular rate error dynamics.

Remark. The inner loop sliding mode dynamics in eq. (132) must be much faster than the outer-loop sliding mode dynamics in eq. (118) to preserve sufficient time scale separation between the loops.

Designing a control function v_c that achieves asymptotic stability with a finite reaching time of the $\bar{\sigma}$ dynamics given as

$$\dot{\bar{\sigma}} = \overline{\Psi}_0(.) + \Delta\overline{\Psi}(.) - [I + E(.)]v_c,\tag{133}$$

where

$$\overline{\Psi}_0(.) = \dot{z}_c - f_2(.) + \overline{C}^1\bar{e}, \quad \Delta\overline{\Psi}(.) = -\Delta f_2(.),\tag{134}$$

we obtained a following expression for the control v_c:

$$v_c = \hat{v}_{c_{eq}} + \overline{R} \cdot \overline{\Sigma},\tag{135}$$

where

$$\hat{v}_{c_{eq}} = \overline{\Psi}_0(.), \quad \overline{R} = diag\left\{\frac{\overline{\omega}_1}{2^{\overline{a}_1}}, \frac{\overline{\omega}_2}{2^{\overline{a}_2}}, \frac{\overline{\omega}_3}{2^{\overline{a}_3}}\right\}; \quad \overline{\Sigma} = \left\{\bar{\sigma}_1^{2\overline{a}_1 - 1}, \bar{\sigma}_2^{2\overline{a}_2 - 1}, \bar{\sigma}_3^{2\overline{a}_3 - 1}\right\}^T.\tag{136}$$

Parameters $\overline{\omega}_i$ and $\overline{a}_i$ in eq. (136) can be identified by analogy with parameters ω_i and a_i in eq. (122).

An allocation matrix $B_v(.)$ is identified as a solution to the following quadratic programming problem:

$$J = \delta_c^T Q \delta_c \xrightarrow[B_2(.)\delta_c = v_c]{} \min\tag{137}$$

as follows:

$$B_v(.) = Q^{-1}B_2^T(.)\left[B_2(.)\cdot Q^{-1}B_2^T(.)\right]^{-1}, \qquad Q \in \Re^{m \times m}\tag{138}$$

where Q is a positive definite matrix. A control allocation algorithm developed in the work[33] can be applied as well.

A "Very" Inner Loop Re-configurable Continuous SMC Design Addressing Actuator Limits

At this point we have a vector-deflection command profile δ_c that is generated by an inner loop finite-reaching-time continuous SMC in eqs. (135) and (136). Being executed this profile will achieve a high tracking accuracy of mission angles in presence of additive and multiplicative uncertainties (failures) without their explicit identification. Now we need to execute the vector-deflection command profile δ_c by actuators, taking into account actuator deflection and deflection limits. A mathematical model of actuators is given as follows:

$$\dot{\delta} = -A_\delta(\delta - u), \quad |\delta_i| \le \delta_m, \quad |\dot{\delta}_i| \le \overline{\delta}_m \forall i = \overline{1,m}\tag{139}$$

A problem is to design a control law u achieving

$$\lim_{t \to \infty}|\delta_{ci} - \delta_i| = 0 \quad \forall i = \overline{1,m}\tag{140}$$

and meeting limitations in eq. (139).

Obviously, a plant in eq. (139) is of a square structure. Therefore, and a design technique of a re-configurable continuous SMC with direct boundary layer adaptation, which is developed in the Chapter I of this

report, is used to address the problem in eqs. (139) and (14). A sliding surface that provides desired tracking dynamics is chosen in the first step of the "very" inner loop SMC design. The vector relative degree of eq. (139) is $\bar{r} = [1,1,...1]^T$, $\bar{r} \in \mathfrak{R}^m$, and the sliding surface is designed in a PI format[10-13]. This is

$$s = \eta + C^2 \int_0^t \eta d\tau = 0, \quad s, \eta \in R^m, \tag{141}$$

where $\eta = \{\eta_1, \eta_2, ..., \eta_m\}^T$, $\eta_i = \delta_{ci} - \delta_i$ and $C^2 = diag\{c_i^2\}$ $\forall i = \overline{1,m}$ are chosen to provide the desired, de-coupled deflection tracking dynamics in the sliding mode. An integral term is added to nullify a steady state tracking error due to continuous implementation of SMC.

Remark. The "very" inner loop sliding mode dynamics in eq. (141) must be much faster than the inner loop sliding mode dynamics in eq. (132) to preserve sufficient time scale separation between the loops.

The second step of the inner loop SMC design is to choose a control function u that achieves asymptotic stability of the s dynamics given by the following

$$\dot{s} = \dot{\delta}_c + C^2\eta - \bar{v}, \tag{142}$$

where

$$\bar{v} = A_\delta (u - \delta). \tag{143}$$

The following Lyapunov function candidate is considered for design of a stabilizing SMC

$$V = \frac{1}{2} s^T s > 0. \tag{144}$$

The derivative of the Lyapunov function candidate in eq. (144) is given by the following

$$\dot{V} = s^T \dot{s} = s^T \left[\dot{\delta}_c + C^2\eta - \bar{v} \right]. \tag{145}$$

The equivalent control v_{eq} is given by

$$\bar{v}_{eq} = \dot{\delta}_c + C^2\eta. \tag{146}$$

The following additional discontinuous control term is constructed to provide negative definiteness of the Lyapunov function derivative in eq. (145)

$$\bar{v} = \bar{v}_{eq} + \rho SIGNs, \tag{147}$$

where $\rho = diag\{\rho_i\}$, $\rho_i > 0$, $SIGN s = diag\{sign\, s_i\}$, $\forall i = \overline{1,m}$. It is easy to show that

$$\dot{V} = -s^T (\rho SIGNs) = -\sum_{i=1}^{3} \rho_i |s_i|, \tag{148}$$

and the sliding surface in eq. (141) will be reached in a finite time.

The elements of the control law vector are developed by solving eq. (143) for the actuator command, u, and inserting the SMC from eq.(147)

$$u_i = \delta_i + \frac{1}{a_\delta} \left(\bar{v}_{eq_i} + \rho_i\, sign\, s_i \right) \quad \forall i = \overline{1,m}. \tag{149}$$

A continuous "very" inner loop SMC with a boundary layer is used to avoid control chattering. The discontinuous terms $sign\, s_i$ are substituted by the following continuous terms

$$sat\frac{s_i}{\varepsilon_i} = \begin{cases} 1, & if\, s_i > \varepsilon_i \\ \dfrac{s_i}{\varepsilon_i}, & if\, |s_i| \le \varepsilon_i \\ -1, & if\, s_i < -\varepsilon_i \end{cases} \tag{150}$$

which results in the following continuous "very" inner loop SMC

$$u_i = \delta_i + \frac{1}{a_\delta}\left(\hat{v}_{eq_i} + \rho_i\, sat\frac{s_i}{\varepsilon_i}\right)\quad \forall i = \overline{1,m}. \tag{151}$$

Although the continuous SMC in eq. (151) prevents chattering, the control function does not account for actuator limits. If there is not enough control power to compensate for disturbances and uncertainties due to damage or other modeling errors, actuator deflection and deflection rate limits may be exceeded and controller integrators may windup. Exceeding actuator limits and integrator windup may lead to significant tracking degradation up to loss of stability and/or pilot induced oscillations. The actuator limitations in eq. (139) are directly addressed by adapting the boundary layer thickness. The boundary layer thickness ε_i is chosen to satisfy the following inequalities that guarantee compliance with actuator displacement and rate limits and avoidance of integrator windup:

a) $\quad |s_i| \le \varepsilon_i \quad \forall i = \overline{1,m}$ - integrator anti-windup, $\hfill(152)$

b) $\quad \left|\delta_i + \dfrac{1}{a_\delta}\left(\hat{v}_{eq_i} + \rho_i\dfrac{s_i}{\varepsilon_i}\right)\right| \le \delta_m \quad \forall i = \overline{1,m}$ - actuator deflection limit compliance, $\hfill(153)$

d) $\quad \left|\hat{v}_{eq_i} + \rho_i\dfrac{s_i}{\varepsilon_i}\right| \le \overline{\delta}_m \quad \forall i = \overline{1,m}$ − actuator rate limit compliance. $\hfill(154)$

Integrator windup is avoided and actuator rate and deflection limits are not violated for the following choice of sliding surface boundary layers

$$\varepsilon_i = \max\{\varepsilon_i^1, \varepsilon_i^2, \varepsilon_i^3\} \qquad \forall i = \overline{1,m}. \tag{155}$$

where

$$\varepsilon_i^1 = |s_i| + r_i^1, \qquad\qquad \forall i = \overline{1,m}, \tag{156}$$

$$\varepsilon_i^2 = \frac{|s_i|\rho_i}{a_\delta\delta_m - (a_\delta\delta_i + \hat{v}_{eq_i})sign\, s_i} + r_i^2, \qquad \forall i = \overline{1,m}, \tag{157}$$

$$\varepsilon_i^3 = \frac{|s_i|\rho_i}{\overline{\delta}_m - \hat{v}_{eq_i}sign\, s_i} + r_i^3, \qquad \forall i = \overline{1,3}, \tag{158}$$

and $r_i^1 > 0$, $r_i^2 > 0$, $r_i^3 > 0 \;\forall i = \overline{1,3}$ are selected to avoid chattering[13].

Remark. The proposed re-configuration (adaptation) of the boundary layers is implemented on-line, and therefore accounts for failures and damage without explicit parameter identification.

Remark. If at least one of the following inequalities is satisfied

$$\bar{v}_{eq_i} signs_i > a_\delta \delta_m - a_\delta \delta_i signs_i \ , \quad \left| \hat{v}_{eq_i} \right| > \bar{\delta}_m \cdot \tag{159}$$

only changing the boundary layer thickness according to eqs. (157) and (158) cannot satisfy the eqs. (153) and (154). The following approach may is used to avoid this anomaly.

Use equivalent control estimate $\hat{\bar{v}}_{eq_i}$ instead of $\bar{v}_{eq_i}$ in the continuous SMC eq. (151) and set it to zero: $\hat{\bar{v}}_{eq_i} = 0$. The following revised expressions may be used in place of eqs. (157) and (158)

$$\varepsilon_i^2 = \frac{\left| s_i \right| \rho_i}{a_\delta (\delta_m - \delta_i \text{sign } s_i)} + r_i^2 , \tag{160}$$

$$\varepsilon_i^3 = \frac{\left| s_i \right| \rho_i}{\bar{\delta}_m} + r_i^3 . \tag{161}$$

Example: Tailess Fighter Re-configurable Control

The tailess aircraft is a 65-degree sweep delta wing, single engine, multi-role supersonic fighter with internal weapons carriage. A mathematical model of a tailess jet fighter that was developed under is the Innovative Control Effectors (ICE) program[33,34] is used in this report to demonstrate abilities of a re-configurable continuous SMC to address aerodynamic surface failures and actuator saturation during execution of flight maneuvers. The ICE configuration includes a large suite of conventional and innovative control effectors. The conventional effectors include elevons, pitch flaps, thrust vectoring, and outboard leading edge flaps. The innovative control effectors include spoiler-slot deflectors and all-moving tips. The following linear model is generated at Mach o.4 and 15,000 ft. altitude for demonstration of the re-configurable SMC technique. Note that thrust vectoring, spoiler-slot deflectors and outboard leading edge flaps are assumed out of order and are not included in the model..

$$
\begin{bmatrix} \dot{\alpha} \\ \dot{\beta} \\ \dot{p} \\ \dot{q} \\ \dot{r} \end{bmatrix} =
\begin{bmatrix}
-0.6344 & 0.0027 & 0 & 0.9871 & 0 \\
0 & -0.0038 & 0.1540 & 0 & -0.9876 \\
0 & -8.2125 & -0.7849 & 0 & 0.1171 \\
-0.5971 & 0 & 0 & -0.5099 & 0 \\
0 & -0.8887 & -0.0299 & 0 & -0.0156
\end{bmatrix}
\begin{bmatrix} \alpha \\ \beta \\ p \\ q \\ r \end{bmatrix} +
$$

$$
+ \begin{bmatrix}
b_{11} & -0.0459 & b_{13} & -0.0133 & -0.0133 \\
b_{21} & 0.0047 & 0 & 0.0031 & -0.0031 \\
b_{31} & -3.7830 & 0 & 1.8255 & -1.8255 \\
b_{41} & -2.5115 & b_{43} & -0.9494 & -0.9494 \\
b_{51} & -0.0453 & 0 & -0.2081 & 0.2081
\end{bmatrix}
\begin{bmatrix} \delta_{el} \\ \delta_{er} \\ \delta_{pflap} \\ \delta_{amt_l} \\ \delta_{amt_r} \end{bmatrix}
\tag{162}
$$

Loss of 50% of a left elevator area and 50% of a pitch flap occur at 5.0 sec and is simulate as follows:

$$b_{11} = \begin{cases} -0.0459, & t < 5.0s \\ -0.02295, & t \geq 5.0s \end{cases}, \quad b_{21} = \begin{cases} -0.0047, & t < 5.0s \\ --0.00237, & t \geq 5.0s \end{cases}, \quad b_{31} = \begin{cases} 3.783, & t < 5.0s \\ 1.8915, & t \geq 5.0s \end{cases}$$

$$b_{41} = \begin{cases} -2.5114, & t < 5.0s \\ -1.2557, & t \geq 5.0s \end{cases}, \quad b_{51} = \begin{cases} -0.0453, & t < 5.0s \\ -0.02265, & t \geq 5.0s \end{cases}, \quad b_{13} = \begin{cases} -0.0395, & t < 5.0s \\ -0.01975, & t \geq 5.0s \end{cases} b_{43} = \begin{cases} -1.9042, & t < 5.0s \\ -0.9521, & t \geq 5.0s \end{cases}$$

Five actuators are modeled as follows:

$$\begin{bmatrix} \dot{\delta}_{el} \\ \dot{\delta}_{er} \\ \dot{\delta}_{pflap} \\ \dot{\delta}_{amt_l} \\ \dot{\delta}_{amt_r} \end{bmatrix} = -\begin{bmatrix} 20 & 0 & 0 & 0 & 0 \\ 0 & 20 & 0 & 0 & 0 \\ 0 & 0 & 20 & 0 & 0 \\ 0 & 0 & 0 & 20 & 0 \\ 0 & 0 & 0 & 0 & 20 \end{bmatrix} \left(\begin{bmatrix} \delta_{el} \\ \delta_{er} \\ \delta_{pflap} \\ \delta_{amt_l} \\ \delta_{amt_r} \end{bmatrix} - \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix} \right), \tag{163}$$

$$|\delta_i| \le 0.5\, rad, \ |\dot{\delta}_i| \le 1 rad\,/\,s \quad \forall i = \overline{1,5} \tag{164}$$

Body rate vector is chosen as command vector

$$y = [p, q, r]^T. \tag{165}$$

The uncommanded variables are defined to be angle of attack α and sideslip angle β.

Angular Rate (Outer Loop) Continuous Finite-Reaching-Time SMC Design.

The vector

$$v = [v_1, v_2, v_3]^T = B \cdot [\delta_{el}, \delta_{er}, \delta_{pflap}, \delta_{amt_l}, \delta_{amt_r}]^T \tag{166}$$

is considered as a virtual control to follow given angular rate command profiles $y_c(t) = [p_c(t), q_c(t), r(t)]^T$, where

$$B = \begin{bmatrix} 3.7830 & -3.7830 & 0 & 1.8255 & -1.8255 \\ -2.5114 & -2.5115 & -1.9042 & -0.9494 & -0.9494 \\ 0.0453 & -0.0453 & 0 & -0.2081 & 0.2081 \end{bmatrix} \tag{167}$$

The sliding surface is designed in a format eq. (132) as follows:

$$\bar{\sigma} = \bar{e} + \overline{C^1} \int_0^t e\, d\tau = 0, \qquad \overline{C^1} = diag\{2, 2, 2\}, \qquad \bar{\sigma}, \bar{e} \in R^3, \tag{168}$$

where $\bar{e} = \{\bar{e}_1, \bar{e}_2, \bar{e}_3\}^T$, and $\bar{e}_1 = p_c - p$, $\bar{e}_2 = q_c - q$ $\bar{e}_3 = r_c - r$ are the command angular rate errors.

The finite-reaching time continuous SMC is designed in a format eqs. (135) and (136) as follows:

$$v_c = \hat{v}_{c_{eq}} + \overline{R} \cdot \overline{\Sigma}, \tag{169}$$

where

$$\hat{v}_{c_{eq}} = 0, \quad \overline{R} = diag\{7.5, 3.0, 3.0\}; \quad \overline{\Sigma} = \left\{ \sqrt[3]{\bar{\sigma}_1}, \ \sqrt[3]{\bar{\sigma}_2}, \ \sqrt[3]{\bar{\sigma}_3} \right\}^T. \tag{170}$$

Control Allocation

The designed control law v_c is transformed into the deflection vector-command δ_c

$$\delta_c = B_v v_c, \tag{171}$$

by allocation matrix B_v, which is calculated using eqs. (137) and (138) with $Q = I$ as follows:

$$B_v = \begin{bmatrix} 0.1196 & -0.1392 & 1.0492 \\ -0.1196 & -0.1392 & -1.0492 \\ 0 & -0.1055 & 0 \\ 0.0260 & -0.0526 & -2.1743 \\ -0.0260 & -0.0526 & 2.1743 \end{bmatrix} \qquad (172)$$

The "Very" Inner Re-configurable continuous SMC Design

The re-configurable continuous SMC u is designed to accurately track the deflection command vector δ_c by compensation of the actuator in eq. (163) addressing actuator limits in eq. (164).

The sliding surface is designed in a format eq. (141). This is

$$s = \eta + C^2 \int_0^t \eta d\tau = 0, \quad C^2 = 0, \quad s, \eta \in R^5, \qquad (173)$$

where $\eta = \{\eta_1, \eta_2, \eta_3, \eta_4, \eta_5\}^T$, $\eta_i = \delta_{ci} - \delta_i \quad \forall i = \overline{1,5}$.

The re-configurable continuous SMC is designed in eq. (151) format with boundary layer direct adaptation in accordance with eqs. (155), (156), (160) and (161). This is

$$u_i = \delta_i + \frac{1}{a_\delta}\left(\hat{v}_{eq_i} + \rho_i \, sat \frac{s_i}{\varepsilon_i} \right), \quad \hat{v}_{eq_i} = 0, \ \rho_i = 1, \ a_\delta = 20 \quad \forall i = \overline{1,5}. \qquad (174)$$

The boundary layers are calculated as follows:

$$\varepsilon_i = \max\{\varepsilon_i^1, \varepsilon_i^2, \varepsilon_i^3\} \qquad \forall i = \overline{1,5}, \qquad (175)$$

where

$$\varepsilon_i^1 = \varepsilon_i^3 = |s_i| + 0.005, \qquad \varepsilon_i^2 = \frac{0.05|s_i|}{0.5 - \delta_i \text{sign } s_i} + 0.005 \qquad \forall i = \overline{1,5}. \qquad (176)$$

Simulations. The following flying quality was enforced during simulations

$\dfrac{p_c}{p^*} = \dfrac{1}{s^2 + 2s + 1}$, $\dfrac{q_c}{q^*} = \dfrac{4}{s^2 + 3s + 4}$, $\dfrac{r_c}{r^*} = \dfrac{.81}{s^2 + 1.8s + .81}$. The results of the simulation, that are given in figures 17-28, show that the prescribed maneuver is accomplished with excellent tracking performance before and after damage to the tailess aircraft without violating the actuator limits.

Conclusions

A re-configurable two-loop continuous SMC method is developed and successfully applied to a realistic tailess aircraft flight control problem. Actuator limit compliance is achieved by sliding surface boundary layer direct adaptation and demonstrated through high fidelity simulations of a tailess aircraft subject to major damage.

References

[1]   Chandler, R. P. "Self-Repairing Flight Control System Reliability and Maintainability. Program Executive Overview," *Proceedings of National Aerospace and Electronics Conference*, May 1984, pp. 586-590.

[2]   Adams, R. J., Buffington, J. M., Sparks, A. G., Banda, S. S., Robust Multivariable Flight Control, Springer-Verlag, New York 1994.

[3]   Pachter, M., Chandler, P. R., and Mears, M., "Re-configurable Tracking Control with Saturation," *Journal of Guidance, Control and Dynamics*, Vol. 18, No 5, September-October 1995, pp. 1016-1022.

[4]   Pachter, M., Chandler, P. R., and Mears, M., "Control Reconfiguration with Actuator Rate Saturation," *Proceedings of American Control Conference*, Seattle, Washington, June 1995, pp. 3495-3499.

[5]   Napolitano, M., and Swaim, R., "Redesign of the Feedback Structure Following a Battle Damage and/or A Failure on a Control Surface by Eigenstructure Assignment," *Proceedings of the Guidance, Navigation, and Control Conference,* 1991, pp. 247-254.

[6]   Gao, Z., and Antsaklis, P. J., "Stability of the Pseudo-Inverse Method for Re-configurable Control Systems," *International Journal of Control*, Vol. 53, No. 3, 1991, pp. 717-729.

[7]   Huang, C. Y., and Stengel, R. F., "Reconfigurable Control Using Proportional-Integral Implicit Model Following," *Journal of Guidance, Control, and Dynamics*, Vol. 13, No. 2, March-April 1990, pp. 303-309.

[8]   Huang, C. Y., "Analysis and Simulation of Control Distributor Concept for a Control-Reconfigurable Aircraft," *Proceedings of Guidance, Navigation, and Control Conference*, August 1988, pp. 726-735.

[9]   Ochi, Y., Kanai, K., "Design of Restructurable Flight Control Systems Using Feedback Linearization," *Journal of Guidance, Control and Dynamics*, Vol. 14, No 5, 1991, pp. 903-911.

[10]   Utkin, V. I., Sliding Modes in Control Optimization, Springer-Verlag, Berlin, 1992.

[11]   DeCarlo, R. A., Zak, S. H. and Matthews, G. P., "Variable structure control of nonlinear multivariable systems: a tutorial," *IEEE Proceedings,* Vol. 76, 1988, pp. 212-232.

[12]   Fernandez, B. R., Hedrick, K. J., "Control of multivariable non-linear systems by the sliding mode method," *International Journal of Control*, Vol. 46, 1987, pp. 1019-1040.

[13]   Slotine, J., Weiping, Li., Applied non Linear Control, Prentice Hall, , N J, 1991.

[14]   F. Esfandiary and H. K. Khalil, "Stability Analysis of a Continuous Implementation of Variable Structure Control," *IEEE Transactions on Automatic Control*, Vol. 36, No. 5, pp. 616-619, 1991.

[15]   Y. Shtessel and J. Buffington, "Continuous Sliding Mode Control," *Proceedings of American Control Conference*, Philadelphia, PA, June 24-26, pp. 562-563, 1998

[16]   Y. Shtessel and C. Tournes, "Flight Control Reconfiguration on Sliding Modes," *Proceedings of AIAA Guidance, Navigation, and Control Conference*, paper # AIAA 97-3632, pp. 1288-1298, August 1997.

[17]   Y. Shtessel, J. Buffington, M. Pachter, P. Chandler, and S. Banda, "Reconfigurable Flight Control on Sliding Modes Addressing Actuator Deflection and Deflection Rate Saturation," *Proceedings of AIAA Guidance, Navigation, and Control Conference,* paper # AIAA 98-4112, pp. 127-137, August 1998.

[18]   M. Azam and S. Singh, "Invertibility and Trajectory Control for Nonlinear Maneuvers of Aircraft," *Journal of Guidance, Control, and Dynamic*, Vol. 17, No. 1, pp. 192-200, 1994

[19] F. L. Lewis, C. T. Abdallah, and D. M. Dawson, Control of Robot Manipulators, Macmillan, NY, 1993.

[20] J. Buffington, and Y. Shtessel, "Saturation Protection for Feedback Linearizable Systems using Sliding Mode Theory," *Proceedings of American Control Conference*, pp. 1028-1032, June 1998.

[21] "Military Standard-Flying Qualities of Piloted Vehicles," MIL-STD-1797A, January 1990.

[22] Rundell, A. E., Drakunov, S. V., and DeCarlo R. A. "A Sliding Mode Observer and Controller for Stabilization of Rotational Motion of a Vertical Shaft Magnetic Bearing," *IEEE Transactions on Control Systems Technology*, **CSS-4**, No. 5, pp. 598-608, 1996.

[23] Y. Shtessel, J. McDuffie, M. Jackson, C. Hall, D. Krupp, M. Gallaher, and N. Hendrix, "Sliding Mode Control of the X33 Vehicle in Launch and Re-entry Modes," *Proceedings of AIAA Guidance, Navigation, and Control Conference*, paper # AIAA 98-4414, pp. 1352-1362, August 1998.

[24] M. E. Jackson and Y. B. Shtessel, "Sliding Mode Thermal Control System For Space Station Furnace Facility," *IEEE Transactions on Control System Technology*, Vol. 6, No. 5, pp. 612-622, 1998.

[25] C. Tournes, and Y. Shtessel, "Aircraft Control Using Sliding Mode Control," *Proceedings of AIAA Guidance, Navigation, and Control Conference*, San Diego, CA, July 29-31, paper # AIAA 96-3692, 1996.

[26] Y. Shtessel, "Sliding Mode Control of the Space Nuclear Reactor System," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 34, No. 2, pp. 579-589, 1998.

[27] F. Zhou and D. Fisher, "Continuous Sliding Mode Control," *International Journal of Control*, Vol. 55, No. 2, pp. 313-327, 1992.

[28] A. Baz, "Continuous Sliding Mode Control of Flow-Induced Vibrations," *Shock and Vibration*, Vol. 2, No. 5, pp. 365-372, 1995.

[29] Y. Shtessel and J. Buffington, "Finite-reaching-time continuous sliding mode controller for MIMO nonlinear systems," *Proceedings of the 37th Conference on Decision and Control*, December 16-18, 1998.

[30] S. Bhat and D. Berstein, "Lyapunov Analysis of Finite-Time Differential Equations," *Proceedings of American Control Conference*, Seattle, WA, June 1995, pp. 1831-1832.

[31] M. Zhihong and X. H. Yu, "Terminal Sliding Mode Control of MIMO Linear Systems," *IEEE Transactions on Circuits and Systems-I*, No. 11, 1997, pp. 1065-1070.

[32] J. Bowlus, D. Multhopp, and S. Banda "Challenges and Opportunities in Tailess Aircraft Stability and Control," *Proceedings of AIAA Guidance, Navigation, and Control Conference*, New Orleans, LA, August 11-13, paper # AIAA 97-3830, pp. 1713-1718, 1997.

[33] J. Buffington "Tailess Aircraft Control Allocation," *Proceedings of AIAA Guidance, Navigation, and Control Conference*, New Orleans, LA, August 11-13, paper # AIAA 97-3605, pp. 737-747, 1997.

[34] A. Ngo, W. Reigelsperger, S. Banda, and J. Bessolo, "Multivariable Control Law Design for a Tailess Aircraft," *Proc. of AIAA Guidance, Navigation, and Control Conference*, paper # AIAA 96-3866, 1996.
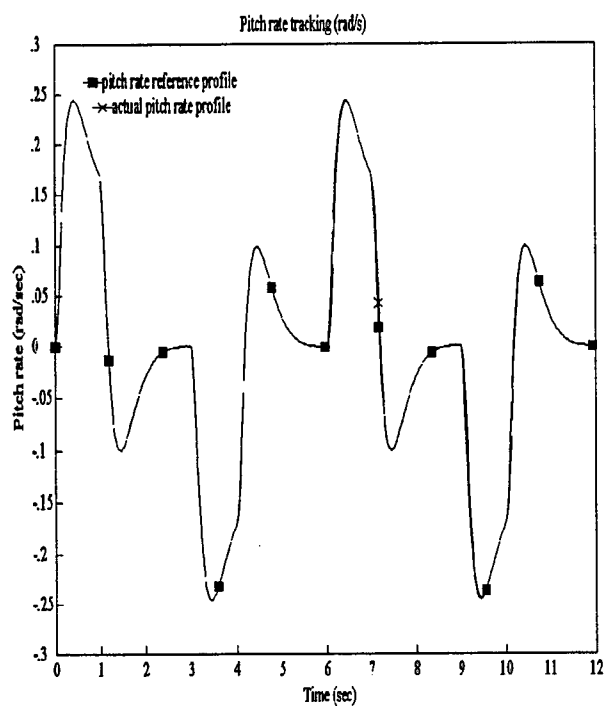
Fig.1   Pitch rate reference profile
tracking via SMC with re-configuration
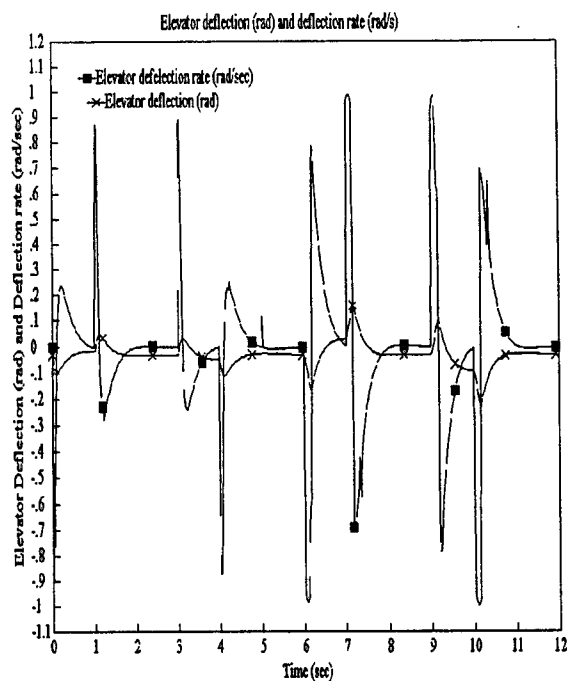
Fig.2   Actuator's deflection and deflection rate
via SMC with re-configuration
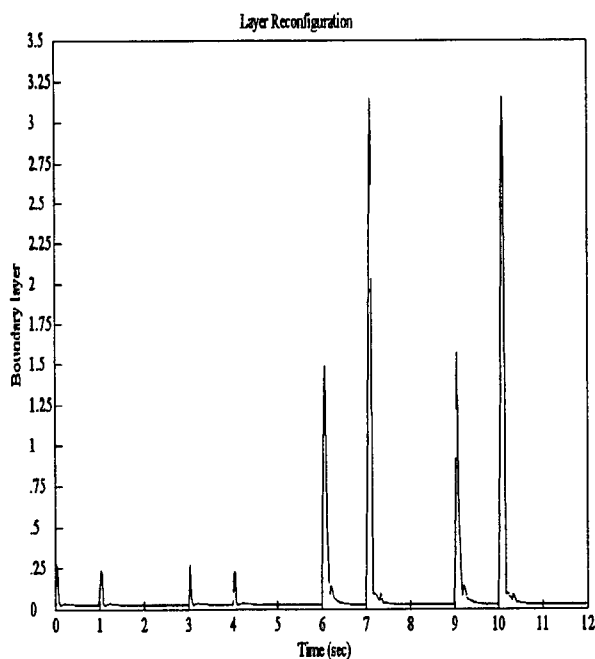
Fig. 3   Boundary layer re-configuration
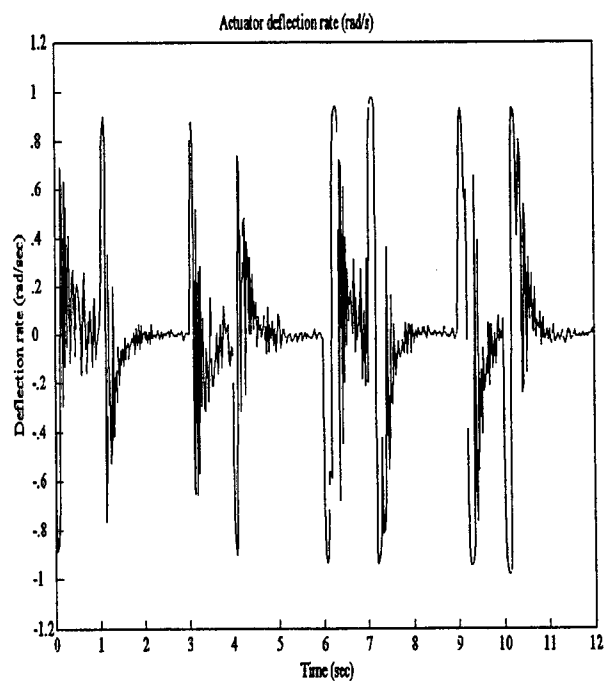
Fig.4   Actuator's deflection rate with
measuring noise and filtered sliding
surface
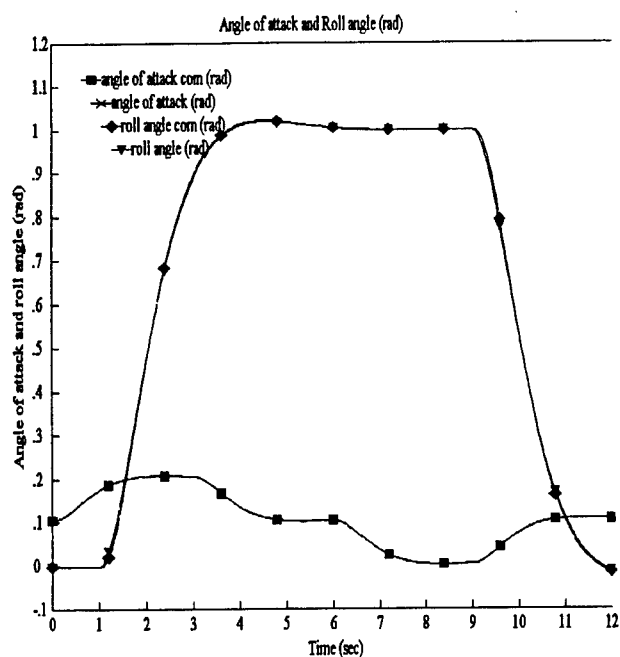
Fig. 5 Tracking of angle of attack and roll angle reference profiles via re-configurable two-loop SMC
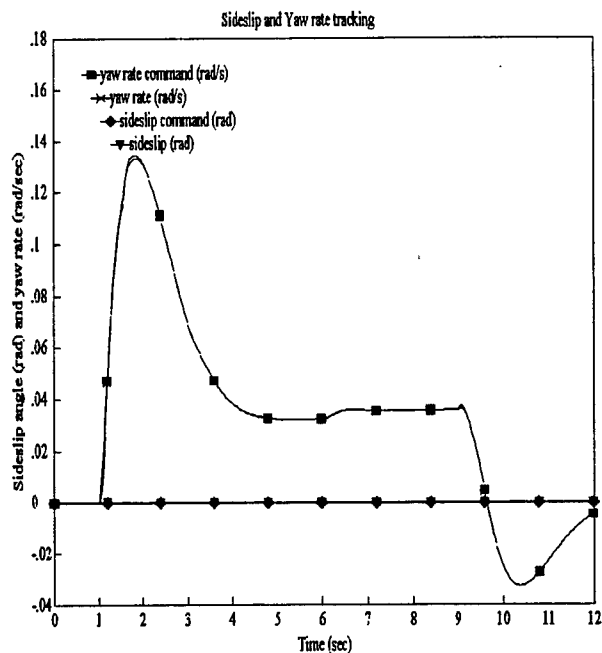

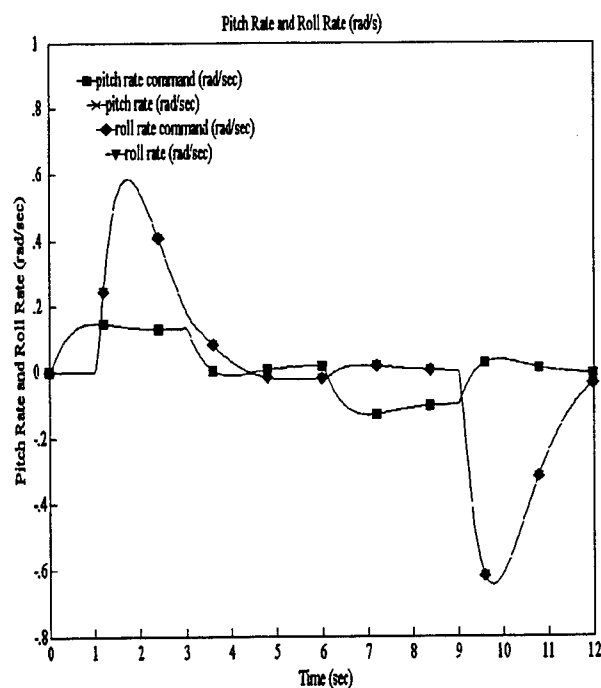**Fig. 6 Tracking of sideslip angle and yaw rate reference profiles via re-configurable two-loop SMC**

Fig. 7 Tracking of pitch rate and roll rate reference profiles via re-configurable two-loop SMC
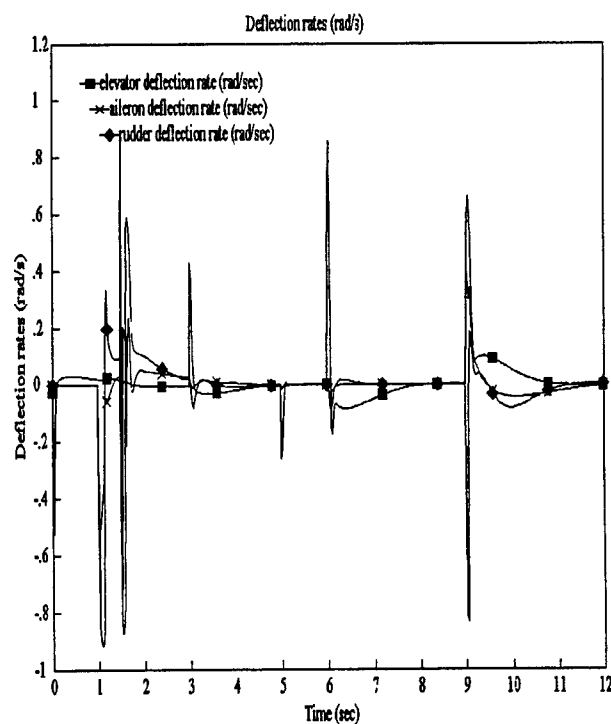
Fig. 8 Aerodynamic surface deflection rates via re-configurable two-loop SMC
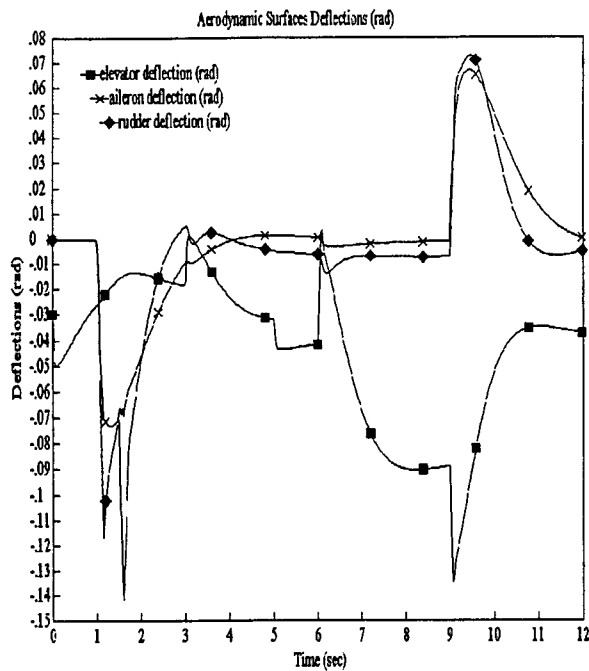
Fig. 9   Deflections of aerodynamic surfaces via re-configurable two-loop SMC
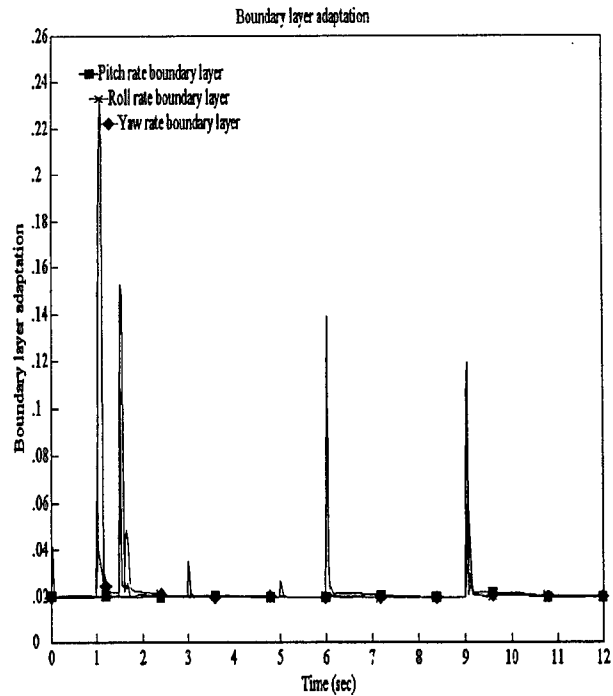
Fig.10   Adaptation of pitch, roll and yaw rate boundary layers of the re-configurable two-loop SMC
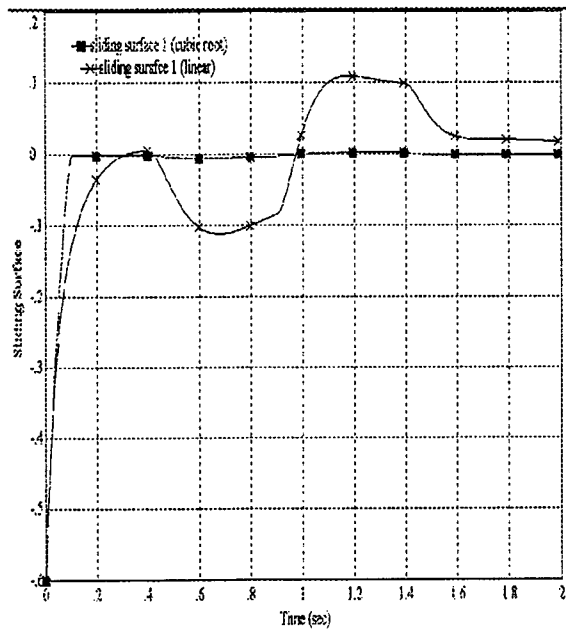
Fig. 11   Sliding surface 1
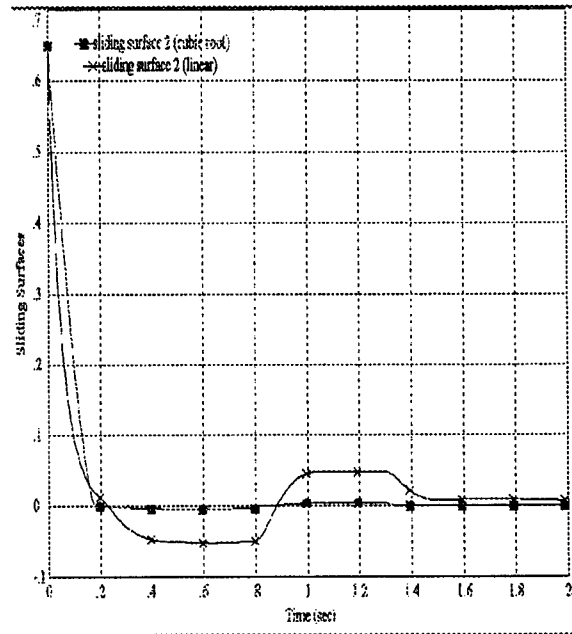
Fig. 12   Sliding surface 2

9-36
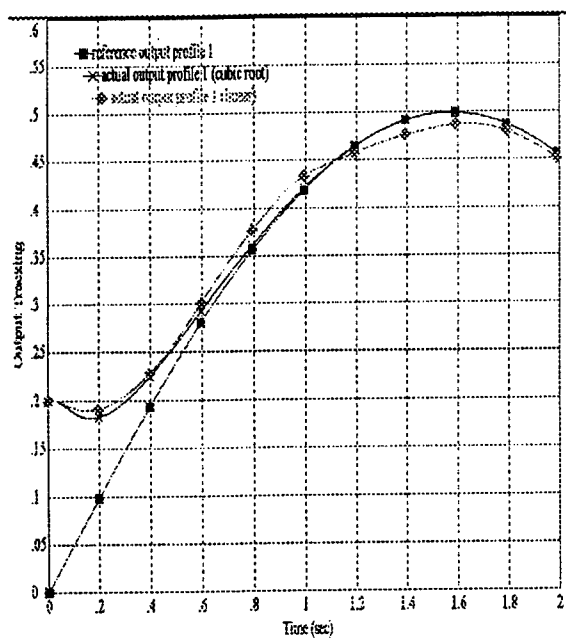
Fig. 13   Output tracking 1
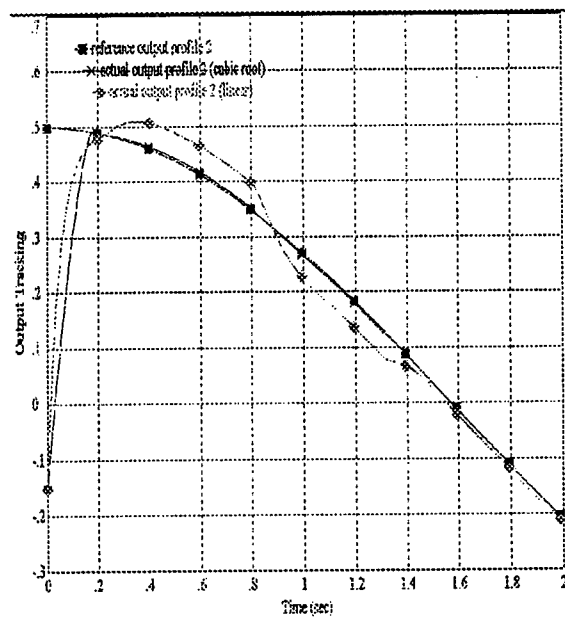
Fig. 14   Output tracking 2
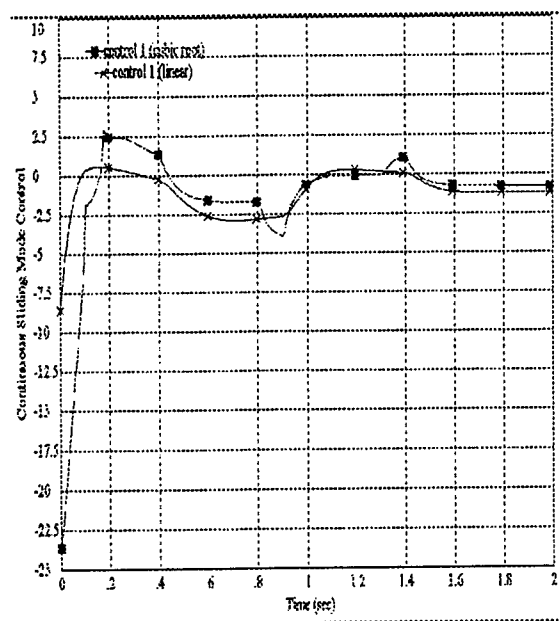
Fig. 15   Continuous sliding mode control 1
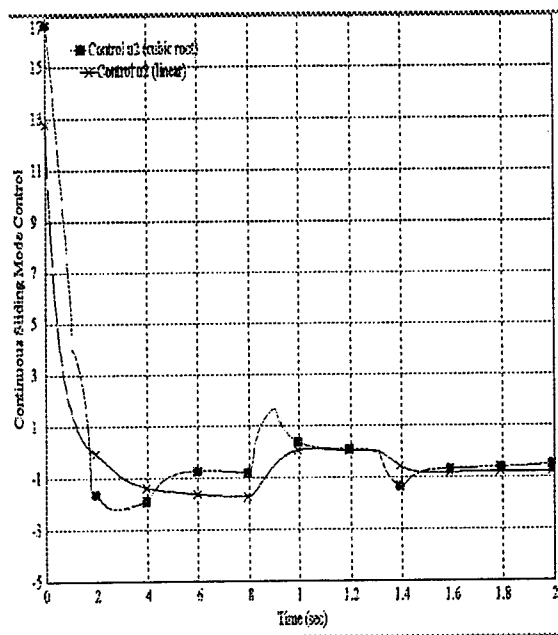
Fig. 16   Continuous sliding mode control 2
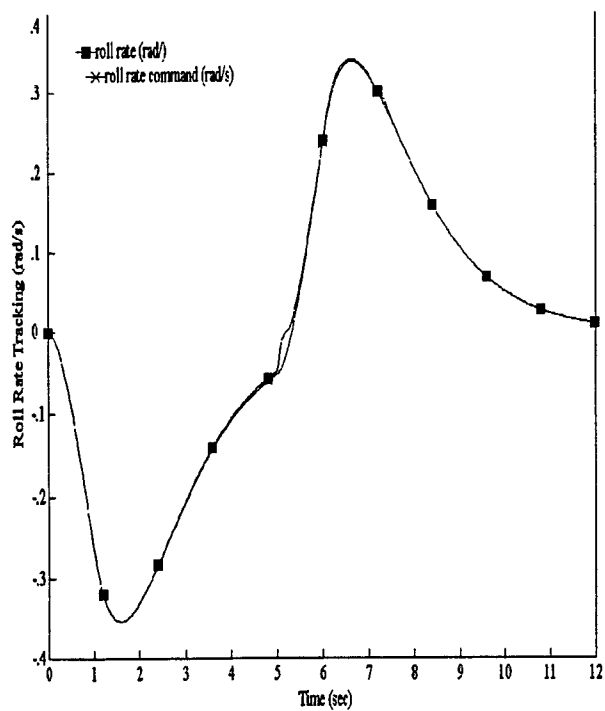
Fig. 17 Roll rate tracking
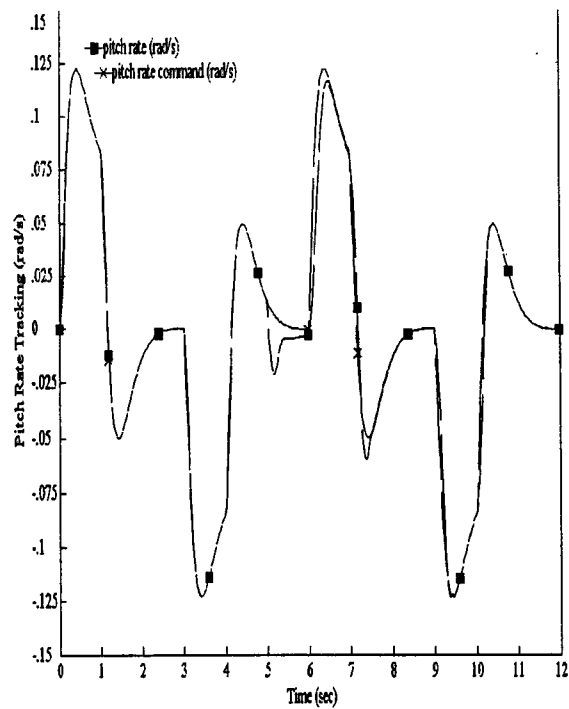
Fig. 18 Pitch rate tracking
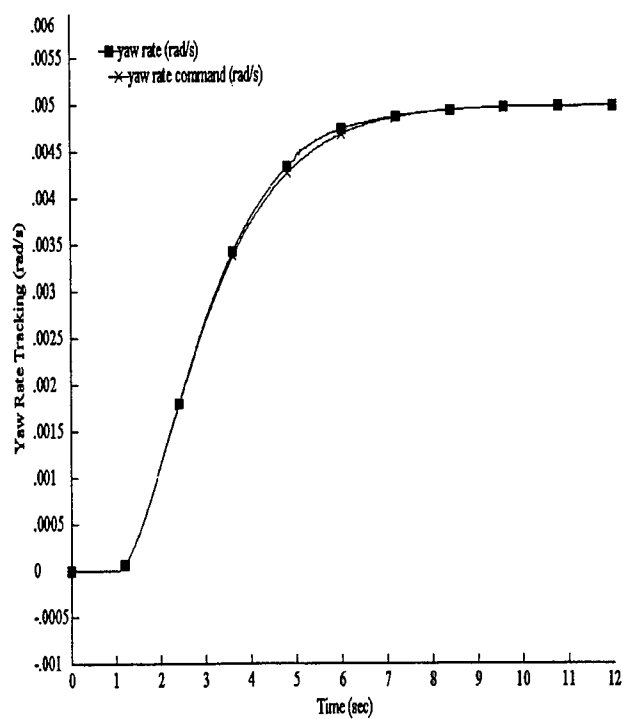
Fig. 19 Yaw rate tracking
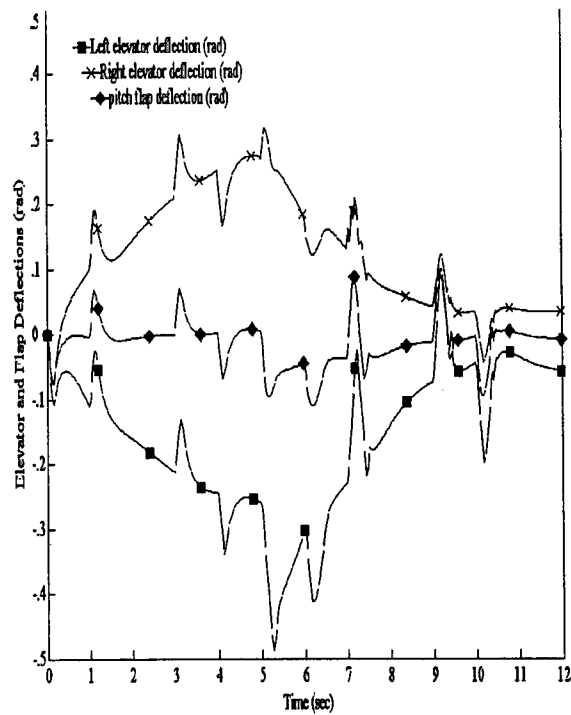
Fig. 20 Elevator and Flap deflections

9-38
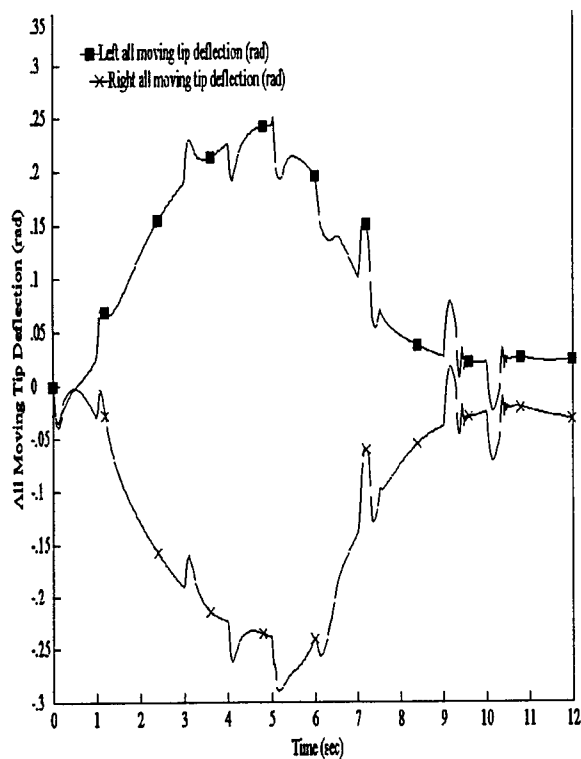
Fig. 21 All moving tip deflections
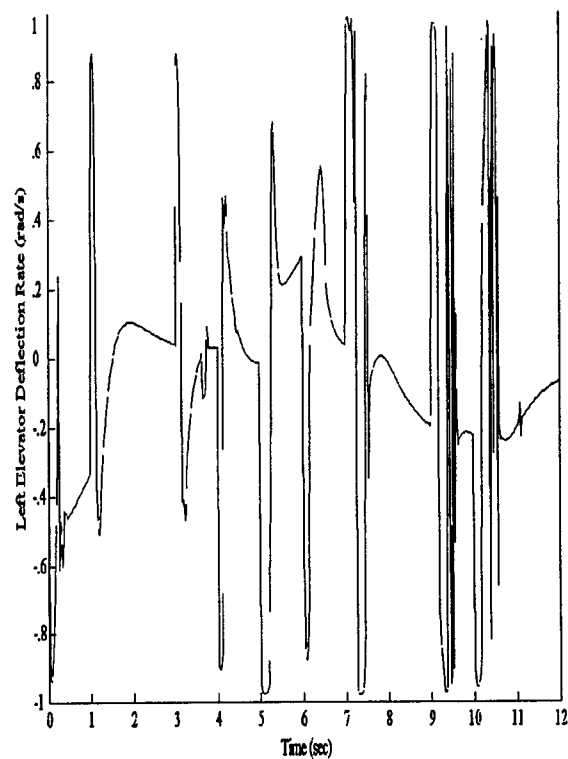
Fig. 22 Left elevator deflection rate
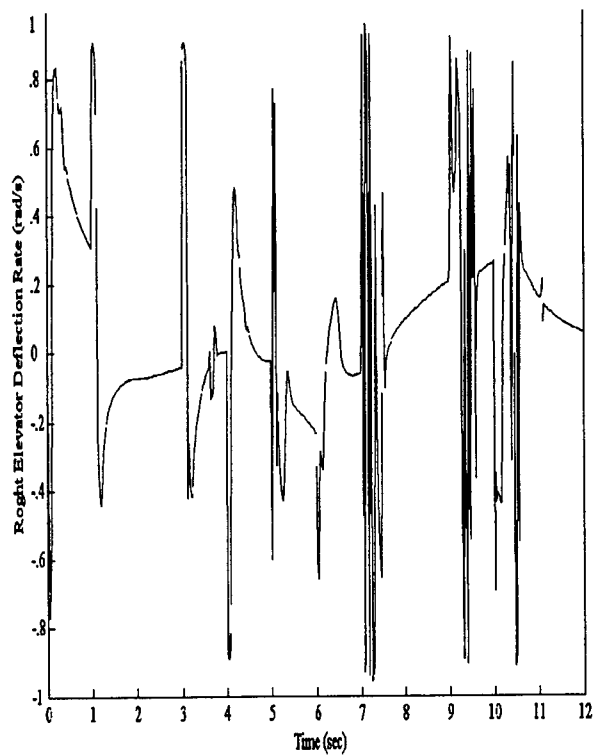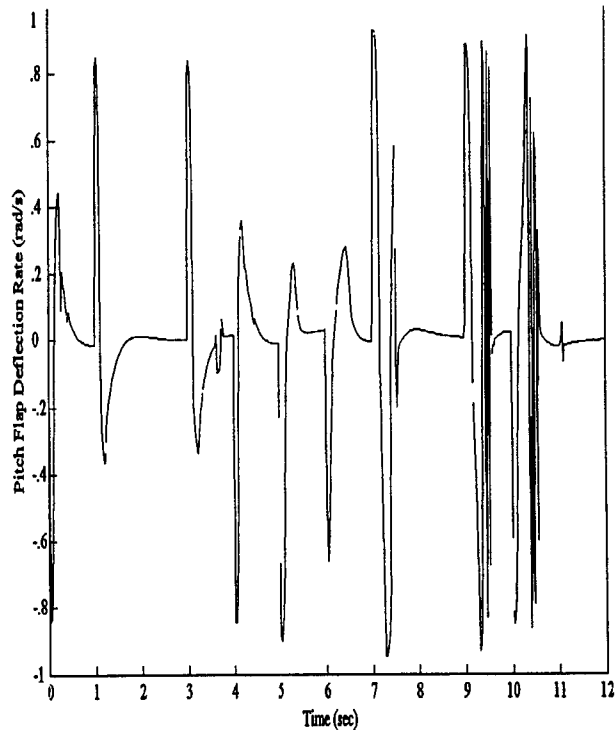

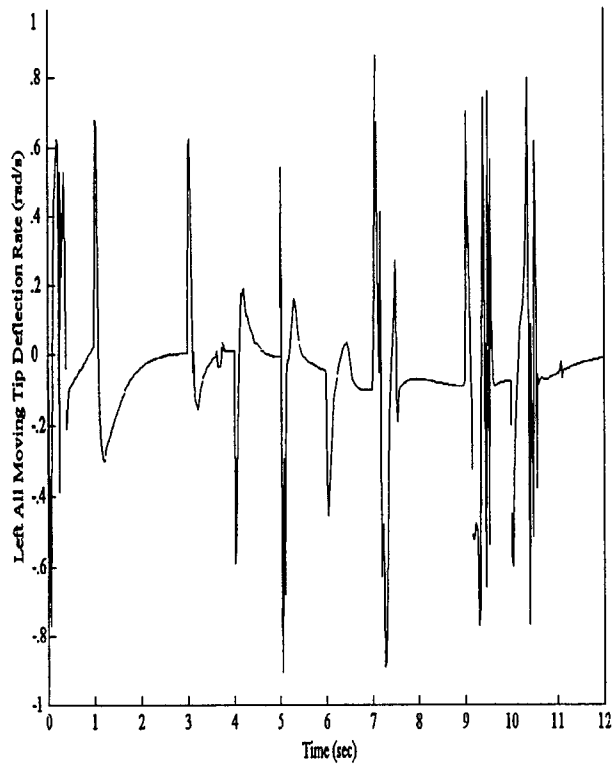Fig. 23 Right elevator deflection rate

Fig. 24 Pitch flap deflection rate

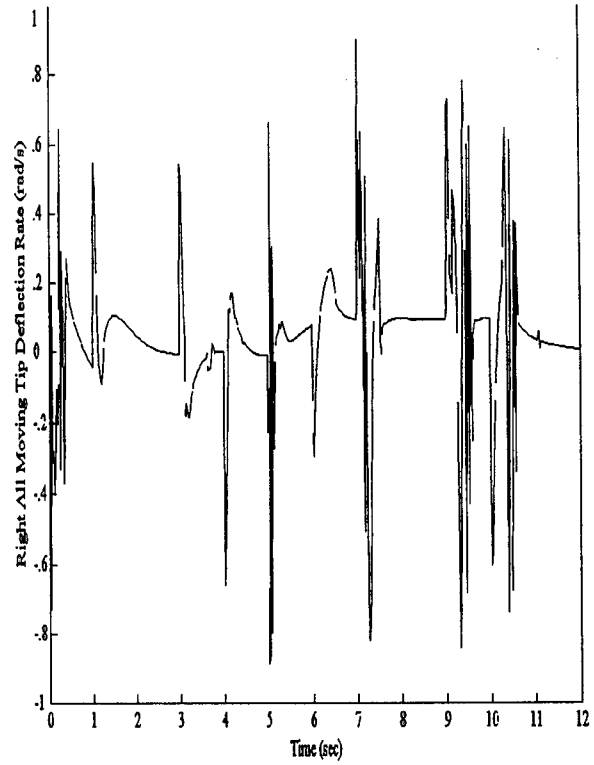**Fig. 25 Left all moving tip deflection rate**

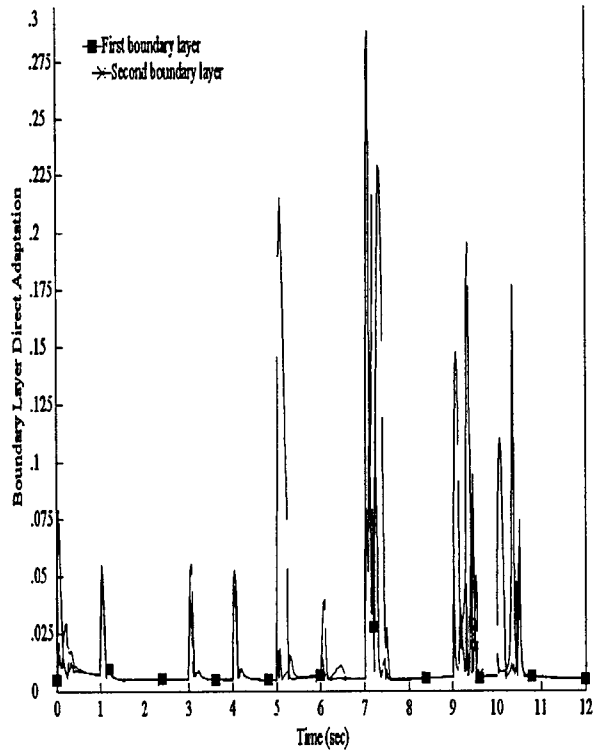**Fig. 26 Right all moving tip deflection rate**

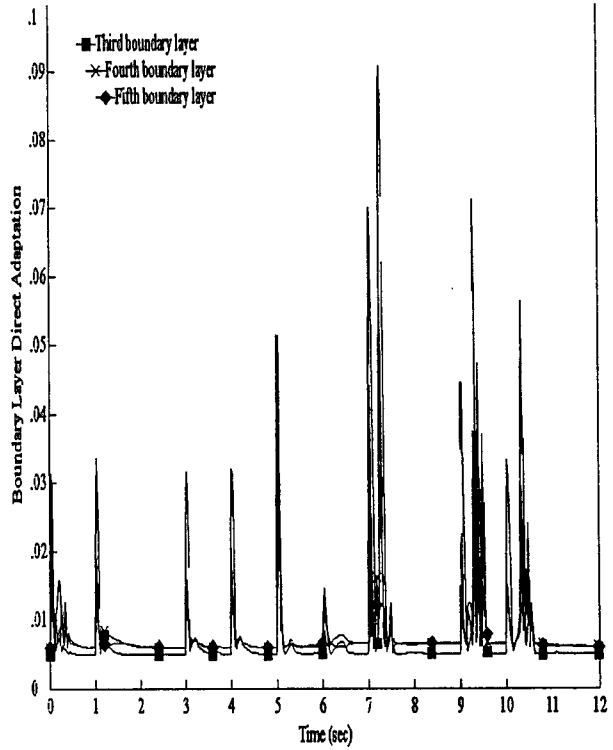**Fig. 27 First and second boundary layers**

Fig. 28 Third, fourth and fifth boundary layers

FEATURE EXTRACTION FOR AIR TO GROUND HRR ATR BASED ON MUTUAL INFORMATION AND STATISTICAL TECHNIQUES

Advanced Approaches

for Classification of HRR Signals

Dr. Janusz A. Starzyk
Professor
School of Electrical Engineering
and Computer Science

Ohio University
Athens, OH 45701-29979

Final Report for
Summer Research Extension Program
Wright Laboratory

January, 2000

Introduction

This report describes statistical methods for feature extraction and classification of air to ground HRR signals. This research is developed under the System Oriented HRR Automatic Target Recognition Program (SHARP). The SHARP objective is to design, evaluate, test, and demonstrate air-to-ground ATR. Target classification is one of the important tasks in target tracking, detection, and recognition within a battlefield scenario. SAR images of moving targets are based on the Doppler effect and quite often are not as clear as the images of stationary targets. Variable terrain and variances of the observed objects make this task more difficult.

A detailed overview of moving target data collection and target modeling issues are presented in [Wes98]. In this research the Moving Target Acquisition and Recognition program data collections 1 & 2, scene 1 were used. These data were produced using X-band 1x1 foot resolution SAR images recorded at 15° and 17° depression angles with 360° coverage in aspect angle. Each 17° data consists of approximately 250 SAR chips per target and 15° data of approximately 195 SAR chips per target. Data collected at 15° were used for testing and 17° for training. 10 target classes were considered with two target classes (BMP2 and T72) having several variants.

The presentation includes results of research on hybrid distance measures and piecewise linear template representation for feature extraction and classification. This work includes establishing hybrid distance measures, describing their statistical properties and tests performed on five-class data set to illustrate quality of these measures. A new piecewise linear strategy to represent statistical information of the training data is described. It is complemented with developed Matlab 5.0 programs for data representation, evaluation of computational cost and quality of developed templates and results of test on HRR data for five and ten class problems. Reference to the baseline method and comparative results for classification accuracy are discussed throughout the entire report. The presented approach requires no prior knowledge of data distribution and provides a reliable performance. As shown in the application results given in this report, the developed hybrid measures and piecewise linear representation of the training data bring improvement to classification results in most situations.

often contains very few samples, usually the concerned region is too small to obtain a statistical estimate of normal parameters, such as covariance.

By the statistical experiments, we find that the minimum distance (called MD) between every two samples coming from a uniform distribution has a predictable maximum value, which is in a positive proportion to the average minimum distance (AMD) of the cluster. Moreover, the ratio of MD over AMD decreases with the increase of the data points number (N) and dimensionality (D). When D and N are large, the ratio stays close to specified value. Fig1-1 gives the ratio vs. N with D = 1-4. These curves are used to determine the threshold Thd_Growth for the cluster growth program.
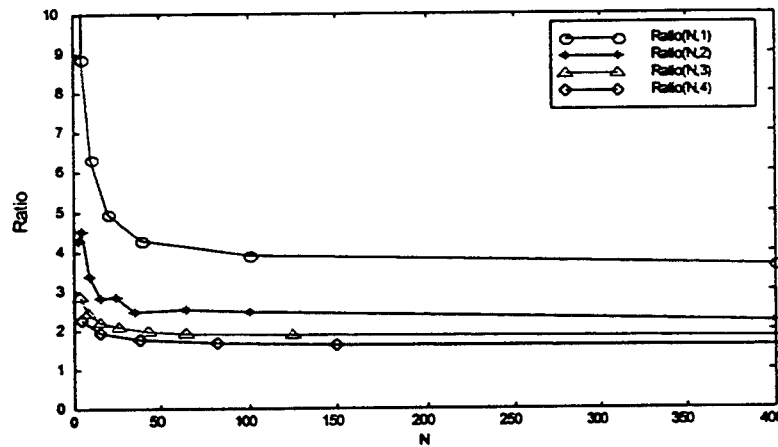

Fig. 1-1 *ratio of MD over AMD vs. N with D=1,2,3,4*

1.1.2 CLUSTER MERGE PROGRAM

After the cluster growth program, we get a lot of small seed clusters, some of which overlap each other. The Cluster Merge program selects the clusters to merge based on a similarity between the pdf (Possibility Density Function) estimates before and after merging.

As shown in Fig.1-2, clusters A and B have their Gaussian distributions $pdfA \sim N(m_A, \Sigma_A)$ and $pdfB \sim N(m_A, \Sigma_A)$. The two distributions form a multi-Gaussian approximation, represented by $pdfW$. Suppose that the points in two clusters are from one Gaussian distribution and should be merged. The distribution after merging is another Gaussian, $pdfM$. The size of the non-overlapped area, represented as NOA, shows the difference of pdfs before and after merging. When NOA is smaller than some threshold, Thd_Merge, the two clusters are merged. The clusters obtained after merging will be checked to merge with other clusters. The cluster merge program repeats checking until there are no two clusters whose distributions are so consistent to be merged.

1.2 Application results

For simplicity, clustering approach is applied on the two-class problem of 2-D data. The data consists of the 2 major features of HRR data extracted by the Singular Value Decomposition (SVD). The classifier is basically a Baysian with the concept of rejection. The performance is compared with the classification without clustering.

In our classifier, the concept of rejection is introduced. As shown in Fig. 1-4, there are 3 parts of the rejection area. Considering the system noise, such as insufficient calculation accuracy and training data, the data points in those areas are rejected without classification. Therefore, there are 3 ratios defined in the output: the misclassification ratio R1, the rejection ratio R2 and the correct classification ratio R3, and their sum equals one.
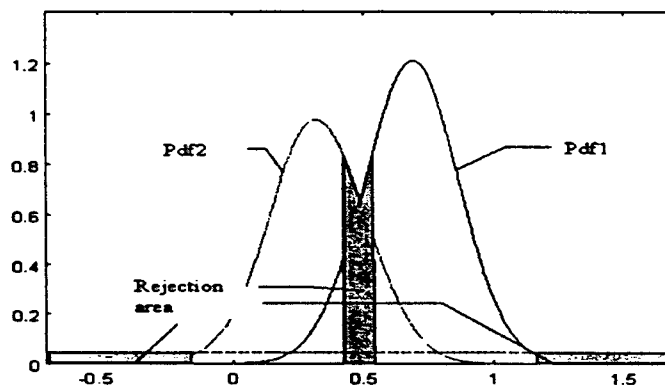

Fig. 1-4 Definition of the rejection area

Fig.1-5 gives the R1 of classifications with and without clustering. The whole experimental results are appended at the end of the report.
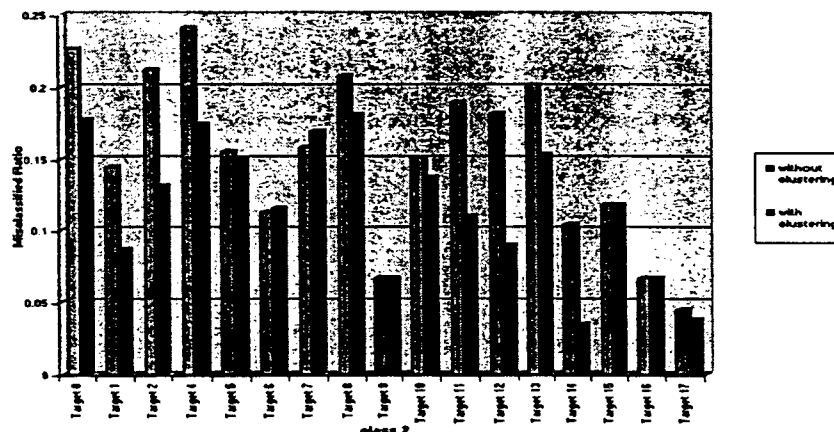

Fig. 1-5 Comparison of R1 with and without clustering
(in the classification between Target 3 and every other target)

3. Clustering increases R2 in most situations. That is, clustering more closely approximates the PDF. The more accurate our approximation of PDF is, the greater a chance we have of finding the area where the local probability densities of two classes are close. This is helpful to our further improvement of classification. Other methods can be applied to the rejected data, and their performance can be accumulated with that of clustering. In this sense, clustering is open to be combined with other approaches.

In spite of the improvement, clustering algorithms has its defects. Since the clustering algorithm approximates the PDF according to the data distribution, naturally, the result is dependent on the training data. When the data is distributed without obvious clusters, our program can not improve the classification result. In some situations, a worse result is obtained because of the noise introduced by the calculation. Clustering algorithm should therefore be applied selectively. Moreover, clustering algorithm is sensitive to changes in the original data. Insufficient data or the difference between the training data and test data can ruin the clustering performance. Therefore, suitable preprocessing is needed before clustering is applied.

2 Hybrid Distance Approach

When observed over the entire aspect angle range HRR signal values vary as shown in Fig. 2-1. This variation requires representation of the training data by many templates. Even though templates may be set as often as one-degree azimuth angle, it still does not represent a perfect fit to the training data. As a result recognition of the test data is difficult and prone to errors.
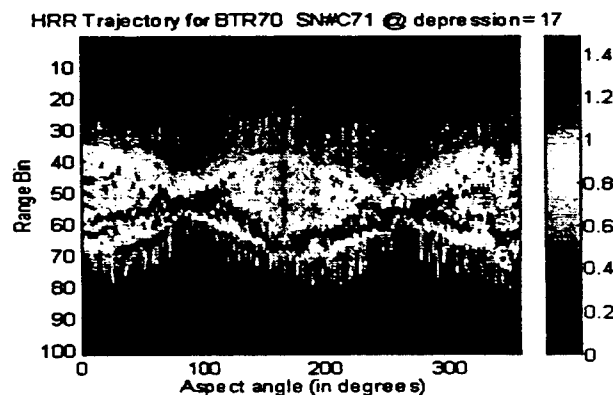

Figure 2-1 HRR return signal strength for various azimuth and range values for BTR70.

One would expect that if the template is capable to correctly represent properties of the test data, then for a template from the same target class the euclidean distance between the template and the observed profile should be minimum. Euclidean distance (ED) is defined as:

represented as MD_D; in the other dimensions, dominated by noise, distance is computed based on the definition of ED, represented as ED_N. HD is the weighted sum of MD_D and ED_N. To define HD, we need to resolve the following problems:

➢ How to divide these two groups of dimensions?

➢ How to adjust the ratio between MD_D and ED_N?

The dominating dimensions can be selected based on the corresponding singular values. When the singular values are smaller than a special threshold value σ_T, we think that the distribution in the related dimension is too concentrated to be considered.

The second problem is more critical. Basically, we want to keep the emphasis on the character of ED_N without losing the adjustment from the MD_D. This is realized by adjusting the singular values to construct a new matrix C_{mtrx} which will replace the inverse of the covariance matrix used in the Mahalanobis Distance. The HD is defined as:

$$HD = \vec{d} \bullet C_{mtrx} \bullet \vec{d}'$$
$$where \quad C_{mtrx} = (V \bullet diag(s_1,\cdots,s_n) \bullet V') \quad\quad\quad (2\text{-}3)$$
$$s_i \text{ is the } \sigma_i^{-2} \text{ after adjustment} \quad 1 \le i \le n$$

Different ways to adjust singular values give us several definitions of HD:

1) HD1 measure

In this definition, we add a noise level to all the dimensions and then normalize all the features by two times the noise level. The noise level is equal to the threshold σ_T, which separates the noise dimensions and major related dimensions. The calculation of C_{mtrx} is described as follows

$$Nl = \sigma_T$$
$$s_i = ((\sigma_i + Nl)/2 \bullet Nl)^{-2} \quad\quad all \text{ the } i, 1 \le i \le m$$
$$C_{mtrx} = V \bullet diag(s_1,\cdots,s_n) \bullet V' \quad\quad\quad\quad\quad (2\text{-}4)$$

2) HD2 measure

As discussed, the singular values are precisely the lengths of the semi-axes of the hyperellipsoid E. The definition of HD2 tries to keep the shape of the hyperellipsoid E but shrinks its volume to a unit size. So all the targets' distributions are normalized and the adjustment from the distributions is based on the relationship between singular values. The matrix C_{mtrx} is obtained as follows

$$\left\{ \begin{array}{ll} \hat{\sigma}_i = \sigma_T & all \text{ the } i, when\ \sigma_i \le \sigma_T \\ \hat{\sigma}_i = \sigma_i & otherwise \end{array} \right.$$
$$s_i = \left(\prod_{l=1}^{m}\hat{\sigma}_l\right)^{2/m} \Big/ \hat{\sigma}_i^2 \quad\quad all \text{ the } i, 1 \le i \le m \quad\quad (2\text{-}5)$$
$$C_{mtrx} = V * diag(s_1,\cdots,s_n) * V'$$

In the next experiment we used the 5-point averaged chip data as testing data instead of O_profiles. The result of Experiment 2 is much better than that of Experiment 1, as shown in Table 2-2.

	Baseline		HD2	
	10.2		8	
	89.8		92	
	72.2		82.2	
	17.6		9.8	
	54		55	
	110		1600	

Table 2-2 Result of Experiment 2
(template: a chip data; observation : 5-point average chip data)

However, the time cost of the algorithm increases too much to be accepted. To improve the efficiency of the algorithm, we merge the neighboring templates to construct a larger template C_{max} by using all the chip data within 10° degree azimuth intervals. This reduces the number of templates to be checked for a single observation without shrinking the search space. The experiment result (Table 2-3 – Experiment 3) is better than that of Experiment 1, but worse than that of Experiment 2. At the same time, its cost is between that of Experiment 1 and 2. These three algorithms can be chosen based on the particular requirement of the real problem. (There is another configuration which uses large template covering 10° degree intervals and O_profiles as observations. That experiment is much faster than Baseline approach, but its performance is worse than that of the Baseline approach. Therefore, that result is omitted in this report.)

	Baseline		HD2	
	10.2		6.6	
	89.8		93.4	
	72.2		77.8	
	17.6		15.6	
	54		59	
	110		400	

Table 2-3 Result of Experiment 3
(template: all chip data within 10° degree; observation : 5-point average chip data)

Within a specified range of the azimuth angles, this function can be approximated by a known functional. The simplest form of this functional is the linear function, that is

$$g'(X, azimuth) = \alpha(X) \bullet azimuth$$

where $\alpha(X)$ is a constant determined by the distribution of X (3-3)

Substituting (3-3) into (3-2), we get

$$f'(azimuth) = \alpha(X) \bullet azimuth + c(X)$$ (3-4)

So, instead of constants, we use linear functions to express the distribution of training signatures at specified azimuth angles, as shown by the dashed line in Fig. 3-1. Furthermore, the new template model of one target is represented by a group of lines instead of a group of constant values used in the template approach. We call it a piecewise-linear (PV
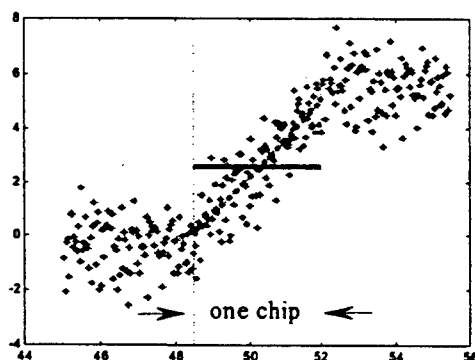

Figure 3-1 Two Template Representations Using the Training Data

The main principle behind the piecewise linear template approximation is to make an explicit use of the training data variation with the azimuth angle by using localized linear regression fits. The corresponding algorithm is described next as the PWL algorithm.

3.1   The PWL algorithm

1.2.2 Linear regression fit to training data

Suppose that we have a vector of i-th range bin values $\vec{X}_i$ corresponding to various azimuth angles, which comprise of the vector $\vec{A}$. The expected multidimensional line projected onto this range bin dimension minimizes the following equation:

$$\left\| \begin{bmatrix} \vec{A} & \vec{I} \end{bmatrix} * \begin{vmatrix} k_i \\ b_i \end{vmatrix} - \vec{X}_i \right\|$$ 3-5)

where $\vec{I} = [1, \cdots, 1]$ R^n

$$Err\_P_1 = \underset{\beta_1,\beta_2,\beta_3 \in R}{minimize} \left\| \vec{P} - \left(\beta_1 \bullet \vec{K} + \beta_2 \bullet \vec{B} + \beta_3 \bullet \vec{I} \right) \right\|$$

$$where \quad \beta_1 = \alpha \bullet \lambda_1; \quad \beta_2 = \lambda_1 \quad \beta_3 = \lambda_2 \tag{3-8}$$

To satisfy these equations we must solve the following:

$$\begin{bmatrix} \left\| \vec{T_1} \right\| & \sum_{i=1}^{n} t_{1i} \bullet t_{2i} & \sum_{i=1}^{n} t_{1i} \\ \sum_{i=1}^{n} t_{1i} \bullet t_{2i} & \left\| \vec{T_2} \right\| & \sum_{i=1}^{n} t_{2i} \\ \sum_{i=1}^{n} t_{1i} & \sum_{i=1}^{n} t_{2i} & n \end{bmatrix} \bullet \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} t_{1i} \bullet p_i \\ \sum_{i=1}^{n} t_{2i} \bullet p_i \\ \sum_{i=1}^{n} p_i \end{bmatrix} \tag{3-9}$$

This calculation is repeated for each shift of the profile vector, and the minimum distance determined for each one, using distance from a profile vector to the multidimensional line segment.

1.3   ROC curves

In this section the confusion matrix and receiver operating characteristic (ROC) curves are used to describe the performance results of the PWL classifier. The ROC curves show the tradeoff between the probability of false alarm P_{fa}, probability of declaration P_{dec}, and the probability of correct identification P_{cc} comprehensively, while the confusion matrix summarizes the performance estimates at a single operating point. We break the three parameters in ROC curves into three sets, represented by three two-dimensional ROC curves. The first ROC curve relates conditional P_{cc} given P_{dec} as a function of P_{dec} and will be refereed to as ROC1. The second ROC curve, referred to as ROC2, relates P_{fa} as a function of P_{dec}. The last ROC curve, referred to as ROC3, shows P_{dec} vs. P_{fa}.

Fig.3-3 and 3-4 summarize the performance of PWL classifier for the 10 target classes, compared to Baseline approach..   In both methods a single-look, with factor = 1.6, was used for MSTAR public targets from collections 1 & 2, scene 1. The overall average results obtained for the Baseline approach can be summarized by P_{id}=78.7 %, P_{dec}=97.2 %, P_{fa}=66 %, while for the PWL method the corresponding results were P_{id}=85.7% (increase by 7%), P_{dec}=96.8% (decrease by 0.4%), P_{fa}=58% (decrease by 8%). PWL method used 10 degrees PWL templates with observations extracted from 5-point averaged chip data. And the ROC curves show that the PWL classifier outperforms the Baseline in most operating points.

4. Conclusion

This report describes a study of statistical methods developed to improve classification performance of HRR based target recognition for air-to-ground images of moving targets. The clustering program described in this report is a parametric approach. It starts with an estimate of the local distribution and efficiently avoids the dependence on having to pre-assume the cluster number. The clustering program is applied to the HRR data classification and improves most experiments. The current application is limited to low dimensionality and a two-class classification problem. In our future work, we plan to extend the clustering close to the real application. To extend the clustering program to higher dimensionality, more experiments on Thd_growth are needed. In addition, designing a multi-class classifier to make use of the clustering results is another important topic.

The piecewise linear method presented uses a localized linear-regression fit to the training data. It uses signal alignment and normalization which minimize least square error between the observation signals and PWL templates. PWL intervals can be selected automatically to minimize the linear regression fit error to the training data. The size of PWL intervals was optimized for classification performance. The method shows a noticeable improvement over the Baseline algorithm on the test data.

The HD measures were developed to capture statistically significant features in the signal space and use them to determine similarity between templates and test data. It was demonstrated that HD measures improve classification performance. The measures depend on information noise level that varies with the classification problem and its training data set. This noise level was arbitrarily selected to classify signals in HRR/ATR problem. Even though the noise level was not optimized in our study, HDs demonstrated an improvement performance over the Baseline approach. Using HD measures, we can optimize computational effort by grouping training data into clusters of different sizes. Both the cluster sizes and their location can be optimized to improve the classification performance. This was done only partially (only the cluster sizes were optimized) in our study. Further optimization, which uses our work on clustering should be performed. Both methods (PWL and HD) were studied independently and each one independently showed an improvement over the Baseline algorithm. Therefore, it is recommended that both approaches should be combined to further improve the classification results.

FEATURE EXTRACTION FOR AIR TO GROUND HRR ATR BASED ON MUTUAL INFORMATION AND STATISTICAL TECHNIQUES

Advanced Approaches
for Classification of HRR Signals

Dr. Janusz A. Starzyk

School of Electrical Engineering
and Computer Science
Ohio University
Athens, OH 45701-29979

Abstract:

This report presents results of research on using clustering approach, hybrid distance measures and piecewise linear template representation for feature extraction and classification of air to ground HRR signals in a statistical approach. This HRR ATR is developed under the System Oriented HRR Automatic Target Recognition, or SHARP program. It includes establishing a statistically supported clustering strategy, developing Matlab 5.0 programs for cluster growth and merging, evaluating clustering quality for feature extraction and classification of the HRR signals, and demonstrating classification accuracy. Next, statistical data representation based on hybrid distance measures was developed, and tests performed on five-class data set to illustrate properties of these measures. Finally, a new piecewise linear strategy to represent statistical information of the training data is described. Both approaches were complemented with developed Matlab 5.0 programs for data representation, evaluation of computational cost and quality of developed templates and results of test on HRR data for five and ten class problems. Reference to the baseline method and comparative results for classification accuracy are discussed following the description of each approach.

1. Clustering

The clustering problem is defined as a problem of classifying a group of data points into a number of clusters without any prior knowledge. According to [Fun 90], the existing clustering methods can be divided into parametric and nonparametric approaches. The parametric approach always has a better performance than the nonparametric one. However, most parametric techniques require an assumption of cluster number, which is not easy to estimate in the real applications.

In this report, we derive a parametric clustering approach to express data distributions by the Multi-Gaussian method. It does not need to pre-assume the number of clusters, making it more suitable for our application.

1.1   DISCRIPTION OF CLUSTERING PROGRAM

To avoid the dependence on the pre-assumed cluster number, our clustering program starts with the estimate of the local distribution. The small clusters, called seed clusters, are constructed by the cluster growth program according to the local distribution of training data. Then the cluster merge program merges those clusters whose distributions are consistent. The merging is performed in order to reduce the complexity of data representation as well as to provide a statistically supported generalization ability for classification.

1.1.1   CLUSTER GROWTH PROGRAM

In the Cluster Growth program, we construct seed clusters in the following way. First, we randomly choose one signal to be the beginning of the cluster, and then absorb the nearest neighbors of the cluster selectively. Here the nearest neighbor of the cluster is defined as the point whose distance to the cluster, or to the nearest point in this cluster, is shortest. This distance will be compared with the threshold, Thd_Growth. If it is within Thd_Growth, this point will be absorbed and the process will be repeated on the next nearest neighbor; otherwise, the cluster growth is terminated.

The Thd_Growth is selected based on the results of the statistical experiments. In these experiments, we assume that data samples in a local neighborhood follow a uniform, rather than Gaussian distribution. That is because a cluster constructed by the seed cluster growth program
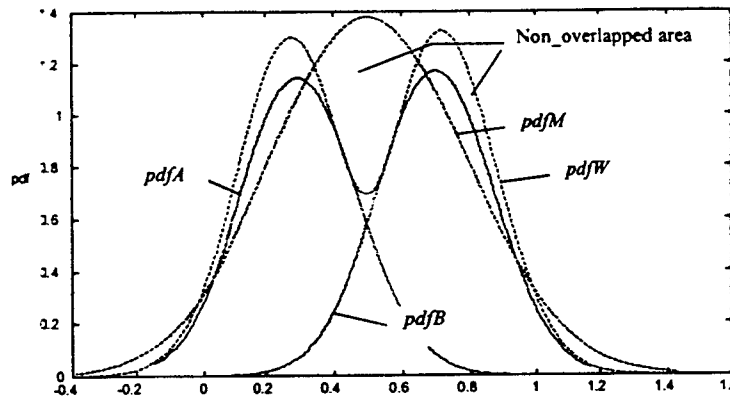
Fig. 1-2 An illustration demonstrating the non-overlapped area

The threshold is determined by the statistic experiments. In the experiments, we use two groups of uniform distributed data to simulate the seed clusters. This assumption doesn't undermine the generality of experimental results. Moreover it is straightward to find whether the two uniform models should merge or not.

The experimental results show the NOA decreases when the means of two clusters get close. And this regularity vanishes when the sizes of the two clusters are much different, as shown in Fig.1-3. In Fig. R_size is defined as the ratio of two cluster's sizes.
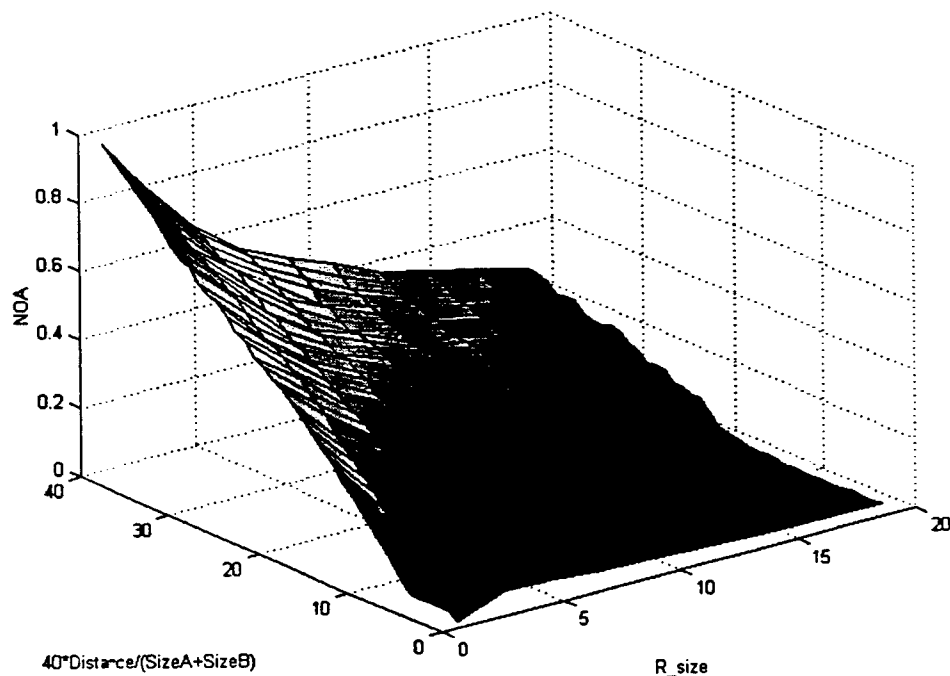

Fig. 1-3 NOA as a function of Distance and R_size

As seen in the results, clustering improves most situations. It helps in the following ways:

1. Clustering can follow the structure of data closely and give a good approximate pdf. It divides the data points into clusters according to the valley existing in the density function without prior knowledge of data distribution. In other words, this parametric clustering method has the same performance as the nonparametric approach. Fig.1-6 shows one clustering result of Target 3 (when classified with Target 2)

2. Moreover, clustering is helpful when dealing with overlapped classes. Fig.1-7 shows the overlapped part of Target 3 and Target 12 with clustering results. In the figure, the points from Target 3 are presented as dots, and their clusters are shown with the dashed ellipses. Five clusters in Target 12 are shown with different symbols (such as circle, star and etc.). As seen in this figure, clustering separates points based on their local distribution. This way, some points from the non-overwhelming class still have a chance to be classified correctly.
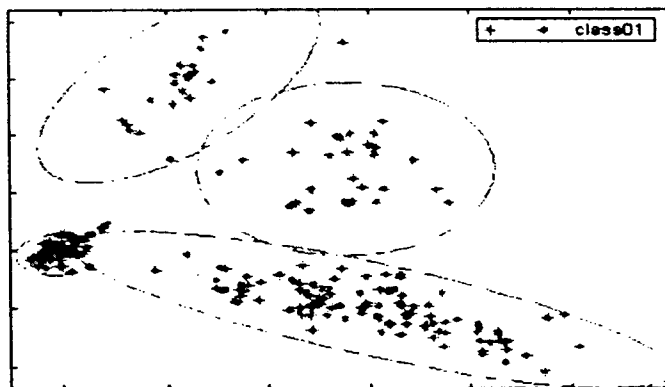
Fig. 1-6 A clustering result of Target 3
(when classified with Target 2)
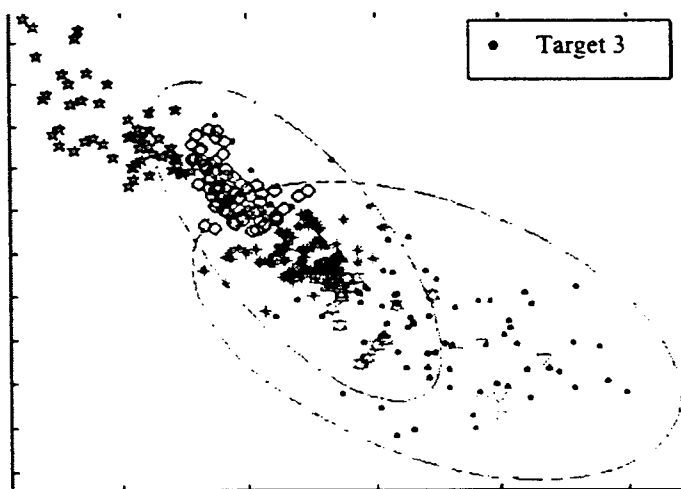
Fig. 1-7 A clustering result on an overlapped area
(overlapped part of Target 3 and Target 12)

10-14

$$ED = \vec{d} \bullet \vec{d}'$$

where $\quad \vec{d} = \vec{O} - \vec{T}$ $\qquad\qquad$ (2-1)

$\vec{O}$ is the observation signature to classify

$\vec{T}$ is the mean of the group of template signatures

When this is not the case the signal can be misclassified. In addition, if one-degree templates represent a good fit to the training data, one would expect that the least square error will be almost constant for the profiles near the template azimuth angle. As illustrated in Fig. 2-2 this is not the case. This figure shows a large variation of the aligned least square error between of 8-point moving average of a single chip data and a template.

To improve our classification, we consider using Mahalanobis Distance (MD) instead. The definition of MD involves the covariance matrix of the template data group:

$$MD = \vec{d} \bullet \Sigma^{-1} \bullet \vec{d}' \qquad\qquad (2-2)$$

where $\quad \Sigma$ is the covariance of the group of template signatures

With the help of the statistical information of template distribution, MD is expected to provide better classification result than ED does.
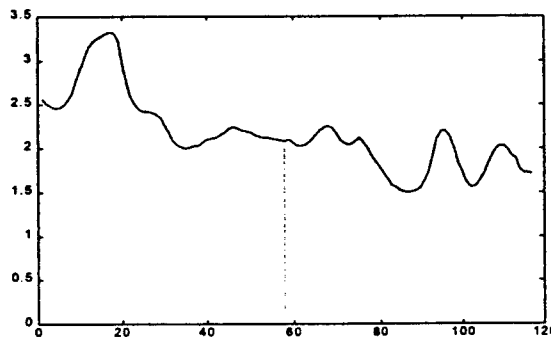

Figure 2-2   **Aligned least square error between of 8-point moving average**

However, MD is not suitable for HRR data. First, its performance is based on the estimate of the statistical characters, therefore the requirements for the training data are strict. Second, it just provides a resolution for the feature extraction in the linear transformation space of the original features. Third, it can't resolve the problem when the observation and template signatures have a different feature space that is typical in HRR data. To overcome these shortcomings, we developed a mixture of ED and MD, called Hybrid Distance (HD).

2.1   Hybrid distance measure

The hybrid distance (HD) combines the MD and ED in the dominating dimensions. In the dominating dimensions, MD is calculated to make use of the statistical information of the data,

3) HD3 measure

HD3 is defined based on the comparison to ED. As it is known,

$$ED = \vec{d} \bullet \vec{d}'$$
$$= \vec{d} \bullet V \bullet V' \bullet \vec{d}' = \vec{D} \bullet \vec{D}' \qquad\qquad where \quad \vec{D} = \vec{d} \bullet V \qquad\qquad (2\text{-}6)$$
$$= \sum_{i=1}^{n} D_i^2 \qquad\qquad\qquad\qquad\qquad\qquad \vec{D} = [d_1, \cdots, d_n]$$

and HD is

$$HD = \vec{d} \bullet C_{mtrx} \bullet \vec{d}'$$
$$= \vec{d} \bullet V \bullet diag(s_1^{-2}, \cdots, s_n^{-2}) \bullet V' \bullet \vec{d}' \qquad\qquad\qquad (2\text{-}7)$$
$$= \sum_{i=1}^{n} s_i^{-2} \bullet D_i^2$$

To keep the dominating character of ED in the hybrid distance measure, HD3 makes $\sum_{i=1}^{n} s_i^{-2} = n$. The calculation required for this distance is as follows:

$$\left\{ \begin{array}{ll} \hat{\sigma}_i = \sigma_T & \textit{all the i, when } \sigma_i \leq \sigma_T \\ \hat{\sigma}_i = \sigma_i & \textit{otherwise} \end{array} \right. \qquad\qquad (2\text{-}8)$$

$$s_i = \hat{\sigma}_i^{-2} \bullet m \Big/ \left(\sum_{l=1}^{m} \hat{\sigma}_l^{-2} \right) \qquad \textit{all the i}, 1 \leq i \leq m$$

$$C_{mtrx} = V * diag(s_1, \cdots, s_n) * V'$$

2.2 Application results

In this section, we applied HD definitions to HRR data with three different configurations. The first experiment uses one chip data to estimate the template model and use O_profiles as testing data, which is similar with the configuration of the Baseline approach. This experiment result is given in Table 2-1. The last row shows the time cost for each approach. All the HD approaches outperform the Baseline in classification performance with the similar time cost. Among them, HD3 and HD2 are a little bit better than HD1.

Table 2-1 Result of Experiment 1
(template: a chip data; observation : O_profile)

	Baseline		HD2	
	10.2		10.2	
P	89.8		89.8	
P	72.2		74	
P	17.6		15.8	
	54		55	
Time	110		115	

Fig.2-3 gives the Receiver Operating Characteristic curves [Wes98] of Experiment 3 compared with the Baseline performance. As shown in Fig.2-3, all the HD approaches have better $(P_{id}|P_{dec})$ [Wes98]than the Baseline, especially the HD2 and HD3. Their P_{fa} is similar with that of the Baseline for large P_{dec} [Wes98] and worse than the Baseline for lower P_{dec}. Similar problem has been discussed in the study of PWL algorithm. (Section 3).
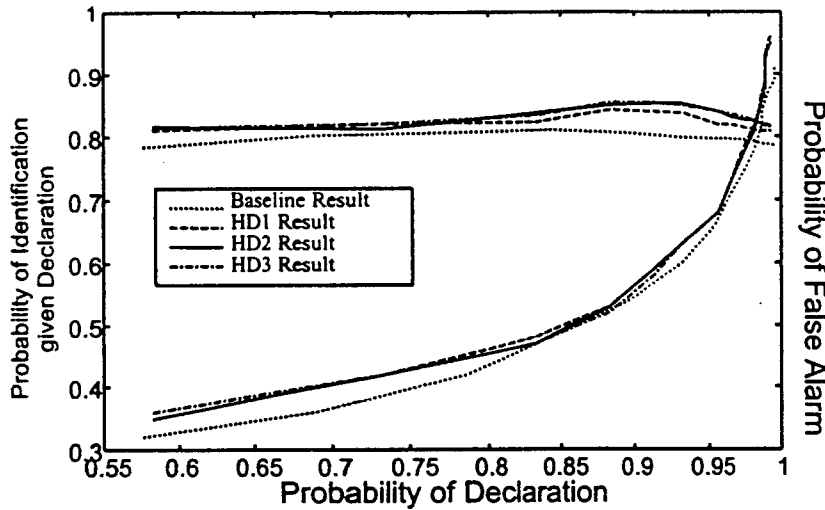

Figure 2-3   **ROC 1&2 of Experiment 3 (HD)**

3   Piecewise Linear Approach

The baseline classification makes use of the mean value of a chip signatures as the template for a specified azimuth and target class. It uses a constant value based on the mean of the chip data at a specified azimuth window.

$$f(azimuth) = mean(X) \ . \tag{3-1}$$

As a result, for each range value, a horizontal line effectively estimates all signatures X in this chip. Even though, data distribution varies with respect to the azimuth, the estimate used in the template approach does not make an explicit use of this variation. If the relationship between the distribution of the training samples and the azimuth value can be estimated and used to construct the training model, the classifier performance would be improved. Suppose that a training data X varies as a function of azimuth as follows:

$$f'(azimuth) = g(X, azimuth) + c(X) \tag{3-2}$$

where $c(X)$ is the constant related only to the training signature X, and $g(X, azimuth)$ is a complicated function and it is expensive and difficult to get its exact estimation.

The solution of this minimization problem yields 2×2 linear equation that provides two linear regression fit coefficients k_i and b_i for each range bin value:

$$\begin{bmatrix} \| \vec{A} \| & \sum_{j=1}^{m} a_j \\ \sum_{j=1}^{m} a_j & m \end{bmatrix} \bullet \begin{bmatrix} k_i \\ b_i \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{m} a_j \bullet x_{ij} \\ \sum_{j=1}^{m} x_{ij} \end{bmatrix} \qquad (3\text{-}6)$$

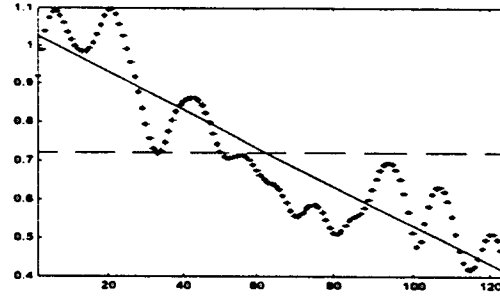Fig. 3-2 gives an example of linear regression fit to the real HRR data of target 2s1b01.

Figure 3-2 Two Template Representations for an example HRR Data
(target 2s1_b01, feature 58 with azimuth =2.2248)

Repeating this calculation for various range bins, we can get two vectors, $\vec{K}$ and $\vec{B}$, which describe the multidimensional linear regression fit to the training data. These vectors are respectively,

$$\vec{K} = [k_1, \cdots, k_n]$$
$$\vec{B} = [b_1, \cdots, b_n]$$

where n is the number of range bin

As a result, training data is described by two constant vectors $\vec{K}$ and $\vec{B}$ rather than a single one used in the template approach.

1.2.3 Profile and template line alignment

The PWL method uses a line equation as a template, which can be expressed by the vectors $\vec{K}, \vec{B}$ on both sides of this line segment as

$$\vec{T}(\alpha) = \alpha \bullet \vec{K} + \vec{B} \qquad \text{for} \quad \alpha_1 \le \alpha \le \alpha_m \qquad (3\text{-}7)$$

The alignment problem is to find the point from this line and the corresponding alignment parameters, λ_1 and λ_2, such that the discriminant function (an error function) of the aligned template and the profile has a minimum norm. For the convenience of calculation, the three alignment parameters are combined and the norm of the error function can be calculated as:
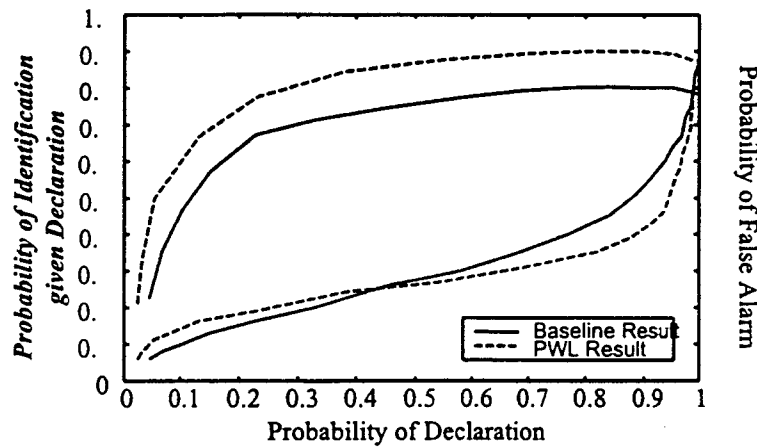
Figure 3-3   **ROC Curves 1 and 2 of PWL Approach for 10 Targets**
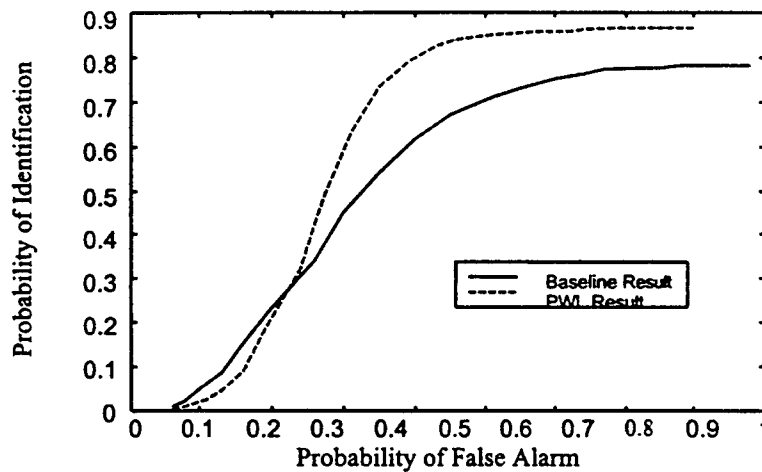

Figure 3-4   **ROC 3 of PWL Approach for 10 Targets**

The PWL classifier with this configuration has a similar computation cost with the Baseline approach but a better performance. The process for 5 targets takes both classifiers about 100 minutes in PC (233MHz). Moreover, we can get a faster algorithm with losing a little performance by adjusting the parameters inside the program, such as the number of observation data. This scalability of PWL is helpful for the future research and improvement of the developed approach.

REFERENCES

[Car95] W. G. Carrara, R. S. Goodman, and R. M. Malewski, *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms*, Artech House, Boston, 1995.

[Irv95] W.W. Irving, L.M. Novak, and A. S. Willsky, *A multiresolution approach to discriminating targets from clutter in SAR imagery*, Proc. SPIE, 1995.

[Fuk90] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, Inc., 1990.

[Mit94] R.A. Mitchell and R. DeWall, *Overview of high range resolution target identification*, Proceedings of the ATR Working Group Systems and Technology Symposium, Monterey, CA, 1994.

[Mit98] R.A. Mitchell and J. Westercamp, *A Statistical Feature Based Classifier for Robust Hight Range Resolution Radar Target Identification*, Submitted to IEEE Trans. on Aerospace and Electronic Systems.

[Mur95] F. Murtagh, "Interpreting the Kohonen self-organizing map using continuity-constrained clustering." *Pattern Recognition Letters* 16, 399-408, 1995. [Sel84] S.Z. Selim and M. A. Ismail, " K-means-type algorithm: generalized convergence theorem and characterization of local optimality", *IEEE Trans. Pattern*

[Sta99] J.A. Starzyk, *Clustering for Classification of HRR Signals*, Progress Report to Air Force Research Laboratory, Sensors Directorate, AFRL/SN, Feb. 1999.

[Wes98] J. Westercamp, *Automatic Target Recognition Using High Range Resolution Radar Signature Features*, Final Report to Air Force Research, Lab Sensors Directorate, AFRL/SN, Sep. 1998.

[Wil97] R. Williams, J. Westercamp, D. Wardell, S. Worrel, and M. Ressler, *Robustness issues for 1D ATR*, Proceedings of the Sixth ATR Working Group Systems and Technology Symposium, Redstone , Alabama, October 1997.